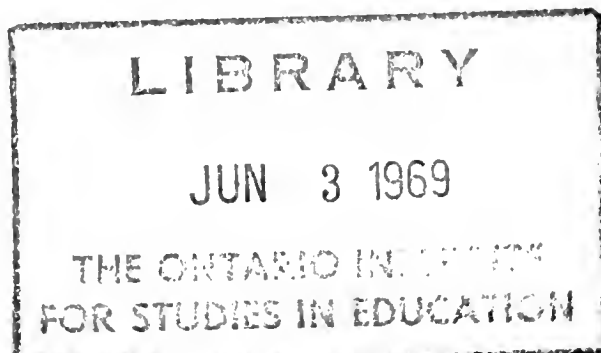


# A STUDY OF FACTORS MEASURED BY THE THORNDIKE INTELLIGENCE EXAMINATION FOR HIGH SCHOOL GRADUATES

BY  
JOHN GRAY PEATMAN, Ph.D.



ARCHIVES OF PSYCHOLOGY

R. S. WOODWORTH, Editor

*No. 128*



NEW YORK

May, 1931

THE LIBRARY

The Ontario Institute  
for Studies in Education

Toronto, Canada



omprises  
Volume  
53-58 ;  
74-78 ;  
ne XVI,  
available

Learning  
LYON.

d Other  
75c.

An Ex-  
ROUSON,

Reaction  
(Cloth,

ousness  
LE IDA

Associa-  
(Cloth,

orizing:  
\$1.25.)  
us Ex-  
THOMAS  
1.10.)  
e-Mark  
ER, JR.

ly Re-  
HAL-

d Rec-  
HILLES.

gence:

College  
\$1.25.  
gments  
OUGH.

Think-

al Se-  
cents.  
Their  
MAR-

actors  
LFRED

Evalu-  
RUTH

nance:

unish-  
T M.

actory  
25.  
curacy  
\$1.25.  
er in  
WADA.

d by  
ATES.

ELSIE

otics:

ty of  
Visual  
HALL.

ility:

A STUDY OF FACTORS  
MEASURED BY THE THORNDIKE  
INTELLIGENCE EXAMINATION  
FOR HIGH SCHOOL GRADUATES

BY  
JOHN GRAY PEATMAN, Ph.D.

ARCHIVES OF PSYCHOLOGY

R. S. WOODWORTH, Editor

*No. 128*



NEW YORK

May, 1931

## ACKNOWLEDGMENTS

To Professor Henry E. Garrett, of the Department of Psychology, Columbia University, I wish to acknowledge my great indebtedness for the suggestion of the problem and for constant guidance and counsel. Without Professor Garrett's cooperation this study could not have been made. To the Department of Psychology, Columbia University, I extend my thanks for financing much of the statistical work necessary for this investigation. This work was done by the Columbia University Statistical Bureau under the direction of Mr. Robert M. Mendenhall. For making available the Thorndike Examination Booklets and the scholastic records requisite to this study I wish to thank both Dr. Ella Woodyard of the Institute of Educational Research, Columbia University, and Mr. Edward B. Fox, Assistant Registrar, Columbia University. To Mrs. Alice Burling Singleton and to my wife, Lillie Burling Peatman, I desire to express my deep appreciation for invaluable assistance.

THE AUTHOR

## TABLE OF CONTENTS

<i>Chapter</i>	<i>Page</i>
I. THE PROBLEM .....	5
II. THE LITERATURE.....	8
III. THE PROCELURE.....	15
A. The Battery of Tests.....	15
B. The Subjects.....	16
C. Treatment of Data for Analysis.....	17
IV. THE RESULTS.....	22
A. Preliminary Analysis of the Thorndike Examina- tion .....	22
B. Intercorrelations of Variables and Reliability Co- efficients .....	23
C. Tetrad Differences of the Five Variables.....	29
D. A Factor Common to the Thorndike Examination	33
E. Relative Importance of the Common and Specific Factors .....	35
F. Relationship of the Test Variables to Scholastic Records .....	40
G. The Probable Nature of the Common Factor. I..	48
H. The Probable Nature of the Common Factor. II.	49
V. SUMMARY AND CONCLUDING INTERPRETATION.....	53
VI. BIBLIOGRAPHY .....	55



# A Study of Factors Measured by the Thorndike Intelligence Examination for High School Graduates

## I. THE PROBLEM<sup>1</sup>

The investigation reported in this paper represents an analysis of the Thorndike Entrance Examination made in the interest of ascertaining what the examination measures and how adequately it measures it. In particular, the investigator has dealt with the question of whether the Thorndike Entrance Examination measures a general intellectual function which possesses some unitary character.

Although the author of the examination has never made any serious claims that this battery of tests does measure, in any adequate sense, a general intellect, or general intelligence, the *use* of the examination entails the assumption, often explicit, that the tests measure *general scholastic ability*. Indeed, some psychologists go so far as to maintain that this examination and others having also a fairly high theoretical reliability coefficient are valid measures of general scholastic ability; that investigations of the validity of the examination are not promising because of the unreliability of validity criteria rather than because of any inadequacy that might accrue to the examination itself.

Thorndike clarifies his own position in this regard in the following statement: "For an ideal examination of the intelligence of candidates for college entrance we might set the following specifications: . . . Significance: The score should correlate as closely with future achievement in college as possible. This maximum possible correlation will not be 1.00, since achievement in college is due in part to health, to freedom from personal worries, and to various moral qualities as well as intellect. Also, the magnitude of the correlation coefficient will depend on the range of the intellect of the candidates, being smaller as that range is restricted. If all the eighteen-year-olds in the country were educated for college, tested, and given a trial in college, we might perhaps expect

---

<sup>1</sup> This study is one of a series of researches (under the general direction of Professor H. E. Garrett) supported by a grant from the Council for Research in the Social Sciences, Columbia University.

a maximum correlation as high as 0.75 to 0.85. Within the restricted range of those who complete a high school course and actually become candidates, we may expect as a maximum 0.55 to 0.65, possibly more." (29).

In the last ten years *future achievement* in terms of college grades has been repeatedly used as the criterion of significance or validity of entrance examinations, giving, on the average, correlations coefficients of 0.40 to 0.50. In a few studies, by selecting the cases so as to eliminate students with *language* handicaps, organic disabilities at the time of the examination, etc., the correlation coefficient has been raised to about 0.65. Even if a correlation of 0.65 rather than 0.40 or 0.50 were taken as the most probable relationship existing between achievement on the entrance examination and achievement in college, approximately only 40 per cent of the variance<sup>2</sup> of college grades could be attributed to factors measured by the examination, and 60 per cent would still be attributable to other factors.

Since the present writer's concern is in regard to the general problem of what the examination measures and how adequately it measures it, reference is made to relationships found to exist between examination and college achievement in the hope that these relationships may help in the determination of the answer to the general problem.

Specific problems considered in this study may be characterized as follows:

1. Are there any functions common to the battery of tests making up the Thorndike Examination? If there are, what per cent of the variance of the total examination may be attributed to common functions, and what per cent to specific functions or other factors?

2. What light do the relationships between the various groups of tests of the battery and the functions attributable to the total examination throw upon the nature of any common and specific or other functions that the examination may measure?

3. What light do the relationships between the various groups of tests of the whole examination and various criteria of college achievement throw upon the nature of any common

---

<sup>2</sup> Variance is defined as the mean of the squares of the deviations from the mean of the distribution; i.e., it is equal to  $\sigma^2$ .

and specific or other functions that the examination may measure?

The answers to these questions will determine the character of the answer to the general question of what the examination measures and how adequately it measures it, as well as to the question: Is the Thorndike Examination a valid or adequate measure of *general scholastic ability*?<sup>3</sup>

---

<sup>3</sup> The writer wishes to emphasize that an individual's Thorndike Entrance Examination score is not the *only* criterion taken into account in determining his fitness for admission to Columbia College. "The psychological examination is simply one part, although a very important one, of the evidence which we consider in passing upon the student's application, but the psychological examination alone has proved to be an exceedingly useful criterion of the student's later performance. Its predictive value is high."—Quoted from a report of a speech made by Dr. A. L. Jones, Director of University Admissions, Columbia University (15).

## II. THE LITERATURE

In his well-known two-factor theory Spearman (26) has maintained that cognitive (roughly, intellectual) functions have in common a general factor, which he characterizes as *g*. The fact that positive correlations are usually obtained between measurements of mental abilities is attributed by him to this common factor. However, these correlations are never equal to 1.00 because, in addition to errors of measurement, each mental function possesses a certain set of specific factors which by definition do not correlate with the specific factors of another function nor with the *g* factor. According to Spearman, then, an individual's score on any mental test, say the Thorndike Examination, would be determined partly by the amount of *g* entering into his mental abilities manifested throughout the examination and partly by factors specific to each particular mental task he performs on the various tests.

Several studies have been made with the aim of analyzing the factors operating in tests making up a *general intelligence* battery. Spearman (26, p. 148 f.) summarizes several such investigations and concludes that "such two independent factors (*g* and *s*) have been demonstrated for at any rate a great number of the sets of tests commonly used for *general intelligence*." He further points out that where the tetrad difference criterion for *g* and *s* factors has not been satisfied, there has been an overlapping of factors between several of the tests because of their great similarity. For example, obtaining the intercorrelations from a battery of seven tests comprising a *general intelligence* test, the criterion was not satisfied due to the overlapping of factors between two completion tests, two analogies tests, and two tests of passages. Removing the overlapping factors by pooling each pair of tests, the criterion for a *g* and *s* factors was satisfied.

It is important to observe that Spearman often adopts an *approximation method* for determining the satisfaction of the tetrad criterion for a *g* and *s* factors. Since the present writer is directly concerned with Spearman's technique and methodology, it seems worth while to cite as an example of the unreliability of Spearman's approximation method the following summary of a study made by Asher.

Asher (2) reports the results of a five-year testing program

for Freshmen at the University of Texas, aiming at the prognostication of success in college. The examination administered the first year (1923) was composed of tests found in ordinary intelligence tests plus some items concerning school subjects. In the 1924 and 1925 examinations more emphasis was given to items pertaining to school information. Each one of the three examinations required fifty minutes of testing time. "In spite of the fact that these tests were loaded with items pertaining to school information, the coefficients of correlation between the scores on the tests and average grades during the first term in college were in the neighborhood of the average coefficients obtained with other intelligence tests." The correlations obtained between the test scores and grades for each of the three years were: .490, 1923; .447, 1924; .438, 1925. In order to find just what the 1925 examination might be measuring and how well the various tests used were measuring it, Asher found the intercorrelations between the tests, between each test and average scholarship, and between each test and each school subject. "These correlations revealed that items pertaining to school information predicted success in their respective school subjects with a comparative high degree of accuracy. The History test predicted history grades very well. The Science test predicted science grades very well. But these tests did not predict grades in other subjects, nor did they predict total scholarship as well as such tests as synonyms, opposites, and reading comprehension. The application of Spearman's tetrad equation to the intercorrelations of the test items revealed much overlapping in the abilities tested by the items. The tetrad equation did not hold throughout the table of correlations."

With these facts at hand (actual coefficients are not reported) Asher constructed a new examination for the 1926 Freshman. It consisted of 100 problems of five different kinds, arranged in ascending order of difficulty in omnibus fashion. All problems except the mathematical ones were of the multiple choice type. The whole examination required forty-five minutes to give and was administered to 805 Arts and Science Freshmen at the beginning of the fall term. He reports the theoretical reliability of the examination as  $.90 \pm .005$ . The five kinds of tests used and their intercorrelations were reported as follows:

TABLE I  
INTERCORRELATIONS REPORTED BY ASHER

<i>Variable</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
1. Opposites		.473	.551	.625	.548
2. Mathematics			.400	.420	.436
3. General Information				.563	.393
4. Synonyms					.582
5. Reading Comprehension					

In order to ascertain whether the abilities measured by this battery of tests could be characterized in terms of a general ability common to the whole examination and specific abilities characteristic of each one of the five tests, Asher applied Spearman's tetrad difference criterion. This criterion involves taking all of the intercorrelations of four or more tests four at a time, thus: taking four variables, 1, 2, 3, and 4, it can be shown that the abilities tested may be thought of as attributable to one general factor plus four specific factors when

$$\begin{aligned} r_{12}r_{34} - r_{13}r_{24} &= 0 \\ r_{12}r_{34} - r_{14}r_{23} &= 0 \\ r_{13}r_{24} - r_{14}r_{23} &= 0 \end{aligned}$$

(26, p. xv; 18, p. 46 f.). Since chance factors are always present, the tetrad differences<sup>4</sup> (right side of the equation) will rarely be exactly zero. The *true* difference, consequently, is estimated from the Probable Error of the obtained difference. If the Probable Error of the obtained difference is large enough to make the difference insignificant, then the proposition is satisfied.

Asher obtained fifteen tetrad differences, the number given by the intercorrelations of five variables. They ranged in size from .080 to .003, the median tetrad difference being .035. The Probable Error (26, p. xi, Formula 16 A) of the differences was equal to .031. To quote Asher, "All differences are within  $\pm$  three times the Probable Error. This indicates that the tetrad equation holds throughout the table of correlations, and that every individual measure of every ability in the table can be divided into two independent parts, *g* and *s*, *g*

<sup>4</sup> Although there is no reason, as pointed out by Pearson and Moul (22), why the term *tetrad* should not be used for brevity's sake instead of *tetrad difference*, the present writer has, on the whole, used the latter term, in accordance with the custom of Spearman, Kelley, and others.

being common to all of the abilities tested. With this fact in mind, it is significant to note that the correlation between the 1926 test and school grades was .605. This fairly high degree of relationship between fall term grades and the combined scores of the five test items whose intercorrelations depend upon the general factor seemingly indicates that scholarship is also dependent in large part upon the general factor." Asher concludes that "it would seem that best results are to be obtained in predicting school achievement with a test in which the abilities measured obey the two-factor theorem."

Asher's tetrad analysis is subject to a criticism, however, which seriously challenges the validity of his interpretation. His Probable Error of the distribution of tetrad differences was obtained by the approximation formula derived and frequently used by Spearman. Although Spearman and Holzinger (27) have recently presented the proof for this formula, as Kelley (18, pp. 13-14) points out, "Spearman assumes that chance would yield a normal distribution of tetrad differences with this standard deviation. This assumption of a normal distribution, even in situations where one general factor only is present does not seem reasonable, for the chance errors in the correlation coefficients are known to be correlated, so that we may expect the chance errors in the tetrads also to be correlated, and to an appreciable extent, for they are only functions of four correlation coefficients and the products of correlation coefficients in pairs are repeated many times in the total population of tetrad differences. This would yield a non-normal distribution, of just what form the writer does not know."

A more reliable estimate of the Probable Error of a tetrad difference may be obtained by the use of the following formula, given by Kelley (18, p. 49, Formula 28) :

$$\begin{aligned} \text{P.E.}_{t_{1234}} = .6745 \frac{1}{\sqrt{N}} & \left[ r_{12}^2 + r_{13}^2 + r_{24}^2 + r_{34}^2 + 2r_{12}r_{14}r_{23}r_{34} \right. \\ & + 2r_{13}r_{14}r_{23}r_{24} - 2r_{12}r_{13}r_{23} - 2r_{12}r_{14}r_{24} - 2r_{13}r_{14}r_{34} - 2r_{23}r_{24}r_{34} \\ & \left. + t_{1234}^2 (r_{12}^2 + r_{13}^2 + r_{14}^2 + r_{23}^2 + r_{24}^2 + r_{34}^2 - 4) \right]^{\frac{1}{2}} \end{aligned}$$

The present writer, using this formula, calculated the Probable Error for the largest of each of the three tetrad differ-

ences of Asher's intercorrelation coefficients, taken four at a time, with the following results:

TABLE II  
TETRAD DIFFERENCES AND PROBABLE ERRORS—ASHER'S DATA

<i>Variables*</i>	$t_{abcd}$	P.E. <sub>t</sub>	$t_{abdc}$	P.E. <sub>t</sub>	$t_{acdb}$	P.E. <sub>t</sub>
1, 2, 3, 4	.035	±.012	.016		—	.018
1, 2, 3, 5	—	.054 ±.012	—	.033		.021
1, 2, 4, 5	.003		.045	±.011		.042
1, 3, 4, 5	.075	±.012	.012		—	.063
2, 3, 4, 5	.067		—	.012		—
						.080 ±.011

\*The names of these tests are given in Table I.

In the case of only one group of tests (1, 2, 3, 4) are all three of the tetrad differences within the limits of four P.E. In each of the other four test combinations, at least one of the tetrad differences is greater than four P.E., in consequence of which, the chances are practically 100 in 100 that the true tetrad differences are greater than zero. Therefore, Asher's conclusion that scholarship is in a large part dependent upon *g* cannot validly be inferred from *these* results.

Wilson (34), working with Professor Burt of the University of London, presents an analysis of an attack on the following problem: "In the discussion of the nature of *general intelligence* and of the possibility of testing it, an important suggestion has lately been made by Thorndike. He had put forward a list of eight tests, based apparently on his own theory, and has implied that they will not satisfy the well-known criteria deduced by Spearman and his collaborators for demonstrating the existence of *general intelligence* as a *central factor*." Wilson gave to a group of "seventy-odd boys" and to a group of fifty girls (ages for each group averaged approximately sixteen years, within a range of 15.5 to 16.5 years) a battery of tests designed to measure, on the basis of Thorndike's suggestion, the following eight tasks:

- |                         |                        |
|-------------------------|------------------------|
| 1. Memory for digits    | 5. Sentence completion |
| 2. Pitch discrimination | 6. Arithmetic problems |
| 3. Opposites            | 7. Number series       |
| 4. Defining words       | 8. Completing pictures |

Thorndike qualified his proposition with the suggestion that the eight tests should have reliability coefficients of .95 and should be given to 10,000 sixteen-year-olds. Wilson had *rela-*

tively few subjects and the reliabilities of his eight groups of tests ranged from .50 to .78.

Analyzing the tetrad differences obtained from the intercorrelations of his battery of tests, Wilson concludes that "it is impossible to do more than say there is a suggestion of the presence of a group<sup>5</sup> factor among the verbal tests, slightly greater suggestion of one between the memory tests, and by far the most evidence is in favor of a group factor between the arithmetic tests."

Garrett (11) reports an analysis of Thorndike's CAVD intelligence examination, given to 338 freshmen girls in Hunter College, Brooklyn, New York. The present writer found the tetrad differences of the intercorrelations of these four tests to be as follows:

$$t_{CAVD} = -.1007 \qquad t_{CADV} = .0189 \qquad t_{CVDA} = .1196$$

Using Kelley's formula, the Probable Error of  $t_{CAVD} = .0227$ .

These tetrad differences apparently satisfy Kelley's XVI Proposition (18, p. 69) that: "If the intercorrelations between four variables are such that  $t_{1234} = t_{1243}$  and  $t_{1342} = 0$ , they could conceivably have arisen from four variables  $x_1, x_2, x_3, x_4$  through which was a general factor plus, in addition thereto, a second factor common to  $x_1$  and  $x_2$  or a second factor common to  $x_3$  and  $x_4$ ." For the above differences, a group factor would possibly be common to C and V, or to A and D. In view of (1) the reliabilities of these four tests; (2) the relatively low correlation between A (arithmetic problems) and the other three tests (C = sentence-completion; V = vocabulary; and D = directions); and (3) Schneck's results (24) according to which, using similar tests, a group factor common to *verbal* tests or to *number* tests, but not general to both, was found,—in view of these three sets of criteria, it might be possible to infer that there is a group factor operating in the CAVD examination common to C and V. Certainly the sentence-completions tests and the vocabulary tests involve a very important factor in common, viz., verbal or linguistic ability. This does not, however, preclude the possibility of a group factor to A and D.

The fact that these results on CAVD very probably do not satisfy Spearman's criterion for *g* and *s* factors certainly lends

<sup>5</sup> When factors overlap several tests but not all tests under consideration, the overlapping function is characterized as a group factor.

additional weight to Garrett's suggestion that other factors, besides verbal ability, are operating here with sufficient strength to permit differentiation *and to render doubtful the meaning and value of a total or composite score.*

Whether the Thorndike Entrance Examination permits the operation of factors, other than verbal ability, that give relatively as important measures of other abilities is one of the problems of this study. This seems to be problematical to the present writer, in spite of the following statement by Thorndike, *et al.* (30): "We measured twenty adults, all high school graduates, chosen from professional and clerical workers, with the Thorndike Intelligence Examination for High School Graduates (average of two forms), and with an incomplete sampling of Intellect CAVD. The correlation is about .95. The self-correlation of the Thorndike Examination score for this group would be only about .975, the correlation of one form with the other being .95. So Intellect CAVD is nearly identical with the ability measured by the Thorndike Examination." That the authors meant this inference from these twenty cases to be taken seriously is certainly to be doubted. Such an intent would be too analogous with the proposition that a measure with a Probable Error of zero is a *true* measure, the Probable Error having been derived from twenty observations that, fortuitously, were identical. (Or, compare Kelley's criticism of Spearman, 18, p. 214.)

In view of the results of the present study to be discussed later, it would appear that the abilities measured by the Thorndike Examination are *not* identical, at least in relative importance, with that measured by Intellect CAVD, unless it be possible that the CAVD examination does not adequately differentiate arithmetic or directions abilities from linguistic or verbal ability.

### III. THE PROCEDURE

#### A. *The Battery of Tests*

The Thorndike Examination given to candidates for admission to Columbia College consists of three booklets of tests, each booklet having alternative forms. The entire examination requires approximately three and a half hours to administer. In order to secure an optimum understanding of the tasks required in the examination, a Practice Form is distributed and answered by the examinees before Booklet I is passed out. The conditions of administration are, on the whole, excellent; particularly, is there the minimum possibility of disturbances and of collusion. Rapport and seriousness are very well guaranteed by virtue of the examinee's acquaintance with the use to which his test score is to be put. The scoring is done by individuals characterized as competent for that work. The author checked a sample of his data for accuracy; no errors were found in those tests scored most "objectively," and practically no disagreement for those test scores in which a certain amount of variability of judgment enters.

The first booklet of the examination consists of nine tests; the second booklet has six; and the third booklet has eight. A characterization of the nature of each test, made on the basis of its construction and the task set, and the time (in minutes) required for each, follows:

TABLE III  
THE TESTS OF THE THORNDIKE ENTRANCE EXAMINATION

No.	Booklet I	Time	Booklet II	Time	Booklet III	Time
1	Following Directions	3	Sentence Completion	8	Reading Comprehension	7
2	Arithmetic	3	Sentence Completion	8	Reading Comprehension	7
3	Arithmetic	6	Sentence Completion	8	Reading Comprehension	7
4	General Information	3	Algebra	10	Reading Comprehension	7
5	Vocabulary	5	General Information	4	Reading Comprehension	8
6	Vocabulary	3	General Information	12	Reading Comprehension	8
7	Vocabulary	7			Reading Comprehension	8
8	Following Directions	7			Reading Comprehension	8
9	Equation Relations	8				

Each one of the Reading Comprehension tests consists of a paragraph of reading matter relevant to various high school studies, which the examinee has to digest in order to answer correctly a series of questions. All of the tests have *speed* as an important common factor.

The examination booklets used in this study were of those candidates for admission to Columbia College who took the tests in June, 1925. This particular group was chosen because, at the time the writer began this investigation, they comprised the most recent class that could have had four years of college work. In all, 568 sets of examination papers from this group of candidates were obtainable.

Three alternative forms of each booklet were used in administering the examination, the ones that a particular individual received being a matter of chance. The forms and the number of students using them were as follows:

Booklet I: Form V,  $N = 304$ ;  
Form X,  $N = 215$ ; Form U,  $N = 49$ .  
Booklet II: Form L,  $N = 270$ ;  
Form N,  $N = 207$ ; Form M,  $N = 91$ .  
Booklet III: Form S,  $N = 287$ ;  
Form J,  $N = 223$ ; Form I,  $N = 58$ .

#### B. *The Subjects*

The group of 568 individuals whose papers were available were homogeneous in at least three respects important to this study: (1) they were of the same sex; (2) they were motivated in the test by the same general purpose, i.e., admission to the College; and (3) they had approximately similar *amounts* of academic nurture (through high school). They were relatively homogeneous in respect to age. Their average age at the time of the examination was probably between seventeen and eighteen years. This average is given as an estimate made on the basis of a sample of 233 men from the total group, since the ages of all 568 men were not available. Of the total group, 233 entered Columbia College in the autumn of 1925. Their average age was 17.4 years,  $\sigma = 1.2$ . Since negative correlations between Thorndike test scores and age were obtained for this group of 233 men, selected in part on the basis of the excellence of their Thorndike scores, the chances are that the average age of the 568 men would be closer to eighteen years than to seventeen, and that the standard deviation would be slightly larger. In terms of absolute age, of course, the group is by no means homogeneous, but *relatively* they tend to be homogeneous in so far as this factor of maturity affects an analysis of this kind, because they represent the upper end of

the maturity level. An age average of ten with  $\sigma = 1.2$  would represent relatively greater heterogeneity.

Other traits that are given by Kelley (18, pp. 24-33) as important in a correlative analysis of mental functions are racial origin and nurture,—training in a general sense, rather than academic nurture only. It is, of course, impossible in view of the present status of knowledge and methodology to render either of these factors perfectly homogeneous with regard to a group of several hundred individuals. Not only are adequate criteria for homogeneous ancestry difficult to establish, but the problem of differentiating individuals according to any possibly adequate criteria is also enormous. Again, in the case of nurture, the problem of measurement in respect to criteria such as social and economic status, cultural and moral background is probably as difficult. Kelley, of course, is emphasizing the fact that these traits should be rendered homogeneous in so far as it is possible to do so. For this group of 568 men there is undoubtedly heterogeneity with respect to these two factors. The influence of such heterogeneity is very problematical. As Schneck (24, p. 11) points out, diversity within a group may raise a correlation, or it may lower it. It would be highly satisfactory to have a group of candidates perfectly homogeneous in these respects. Since the groups of candidates that apply for examination are not homogeneous, and since this study represents an analysis of the Thorndike examination as it is used, such heterogeneity as exists in respect to these traits of ancestry and nurture may be characterized as typical for a group of such candidates.

To recapitulate,—the group is quite large, homogeneous with respect to sex and motivation in the test situation, and very probably quite similar with regard to age and education. It is heterogeneous with respect to ancestry, economic and social status, cultural and moral background,—typically so from the point of view of such a group of men, selectively so from the point of view of the universe.

### *C. Treatment of Data for Analysis*

#### *1. Arrangement of Test Groups*

An analysis of the twenty-three tests comprising the three booklets of the Thorndike examination revealed similarities that could be utilized in setting up a battery of five groups of

tests, such that the tests *within* each group were relatively similar with respect to construction and the task set and such that all of the tests of one group were fairly different in these respects from those of another group. The five groups of tests, with names suggestive of their nature, the time required and the maximum possible score for each are as follows:

1. *Reading Comprehension*: the eight tests of Booklet III; sixty minutes; maximum possible score = 144 (x 2).

2. *Verbal or Linguistic Ability*: the three vocabulary tests of Booklet I and the three completion tests of Booklet II; thirty-nine minutes; maximum possible score = 218.

3. *Number or Mathematical Ability*: the two arithmetic tests of Booklet I, the equation relations tests of Booklet I, and the algebra test of Booklet II; twenty-seven minutes; maximum possible score = 135.

4. *Ability to Follow Directions*: test 1 and 8 of Booklet I; fourteen minutes; maximum possible score = 65.

5. *General Information*: tests 4 of Booklet I, and 5 and 6 of Booklet II, fifteen minutes; maximum possible score = 220.

This arrangement of the twenty-three tests does not rest on the assumption that there is no overlapping of abilities from one group to the other. The question of such overlapping is one of the problems involved in this study. Furthermore, the descriptive terms used to characterize the nature of each group are not assumed to be exact designations of the functions involved. They are so used for convenience in reference and at the same time are undoubtedly meaningful in suggesting differences in content or task from group to group.

Undoubtedly achievement in any of the twenty-three tests requires verbal ability, that is, a knowledge of words and ability to use them. The six tests characterized as the Verbal Ability group appear to involve, predominantly, a knowledge of words and their usage. The Reading Comprehension group, although involving an extensive knowledge of words and their usage, has been differentiated from the Verbal Ability group on the basis of test construction and task set. The correctness of an individual's answer in the Reading Comprehension tests is a function of his assimilation and understanding of the *content* of the paragraphs of prose material given in each one

of the eight tests. The four tests of Number Ability, although requiring some verbal ability, also require a knowledge of number symbols and concepts and the ability to use them.

The time required and the number of tests for each of the fourth and fifth groups are certainly not ideal for a battery of five groups of tests designed to measure five groups of intellectual functions. This examination, however, was not constructed to measure specific abilities well; rather, it was designed to measure a sample of general scholastic ability. The five groups of tests have been set up here for purposes of analysis characterized in the introduction.

## 2. *Reduced Scores*

To correct for slight differences in difficulty among alternative forms of the same booklet, the author of the examination assigned weights to be added or subtracted to the total score of a particular booklet. In dealing with the tests in the five groups, characterized in the preceding section, these weights could no longer be logically assigned, since, in any one of three of the five groups, tests were derived from both Booklets I and II. To avoid constant errors that might be attributable to the lack of adequate weightings for differences in difficulty, the scores for each of the groups were transmuted into T-scores, according to the Form or combination of Forms used by an individual. This entailed setting up fifty T-score tables with theoretical means and standard deviations of 25.0 and 5.0, since each one of the five groups of tests was dealt with in halves, in order to get theoretical reliability coefficients. For example, for each half of the Reading Comprehension group, three T-score tables were necessary, one for 281 individuals using Form S, one for 223 using Form J, and one for 58 using Form I. Six of the 568 cases had to be discarded in setting up the T-score tables, since they represented the total number of scores combined from the U Form of Booklet I and the M Form of Booklet II, obviously too few from which to derive a distribution. The average number of cases used to derive each of the fifty tables was 112.

All intercorrelations and other results involving the five groups of tests are presented from both raw, or original scores, and reduced scores. Because of differences in difficulty between the various Forms, results based on the reduced scores

are probably more reliable and valid, and consequently are given greater weight in the discussion and interpretation.

### 3. *Scholastic Records*

Of the 568 men taking the examination in June, 1925, 233 entered Columbia College the following autumn. These individual's scholastic records were obtained and their letter grades for each course were quantified and weighted according to the number of hours the course convened each week. Not only were these weighted grade scores recorded according to the year or summer session in which they were received, but they were also classified into the following categories: Architecture, Astronomy, Biology, Chemistry, Classics, Contemporary Civilization, Economics, Education, Engineering, English, Fine Arts, General Honors, Geology, German, Government, History, Law, Mathematics, Medicine, Music, Philosophy, Physics, Physical Education, Psychology, Romance Language, and Sociology. A few other titles, occurring very infrequently, were also recorded according to their course name.

In the correlative analysis six groups of grade criteria were set up: (1) Contemporary Civilization; (2) English; (3) Foreign Language; (4) Science and Mathematics; (5) Social Science; and (6) Total Grades. The fourth-year grade scores were not included in these criteria because of the specialized work, such as law, architecture, etc., taken by a large per cent of the individuals in their senior year. In all, 159 of the entering group were available for the six grade criteria, each one having at least six hours work (ordinarily two semesters) in each of the first five categories. The total grade scores were derived from *all* of the grades received during the first three years. For a large majority of individuals, approximately ninety per cent of their total grade scores were represented by the grades of the five divisions.

The following system, similar to the more or less arbitrary ones of other investigators (Cf., 8, and 35), was used in quantifying the letter grades:  $A + = 12$ ;  $A = 11$ ;  $A - = 10$ ;  $B + = 8$ ;  $B = 7$ ;  $B - = 6$ ;  $C + = 5$ ;  $C = 4$ ;  $C - = 3$ ;  $D = 2$ ;  $F = .5$ . In this system the difference between A and B is slightly greater than the difference between B and C in accordance with the probabilities of a corresponding greater difference in the assignment of grades. Similarly, the difference

between D and F is greater than the difference between B and C or C and D.

#### 4. *Recording Data and Computations*

All of the test and grade scores were carefully checked in recording and later rechecked to guarantee a high probability of accuracy, this work being done by Lillie Burling Peatman, Alice Burling Singleton, and the present writer. The T-score tables and the correlation coefficients, with means and standard deviations, were computed by the Columbia University Statistical Bureau under the direction of Robert Mendenhall (32). Computations not made by the Statistical Bureau, such as correlations of sums, tetrad differences and their probable errors, Beta coefficients, etc., were calculated by the writer, using a comptometer and verifying results carefully.

#### IV. THE RESULTS

##### A. *Preliminary Analysis of the Thorndike Examination*

Inasmuch as the group of 159 individuals whose three-year college grades were available is used in the analysis of the factors measured by the examination, there is here presented a rather detailed comparison of this group with those taking the examination in June, 1925 ( $N = 562$ ) and with those of this total group who entered the college in the autumn of that year ( $N = 232$ ). The comparisons are made on the basis of the reduced scores.

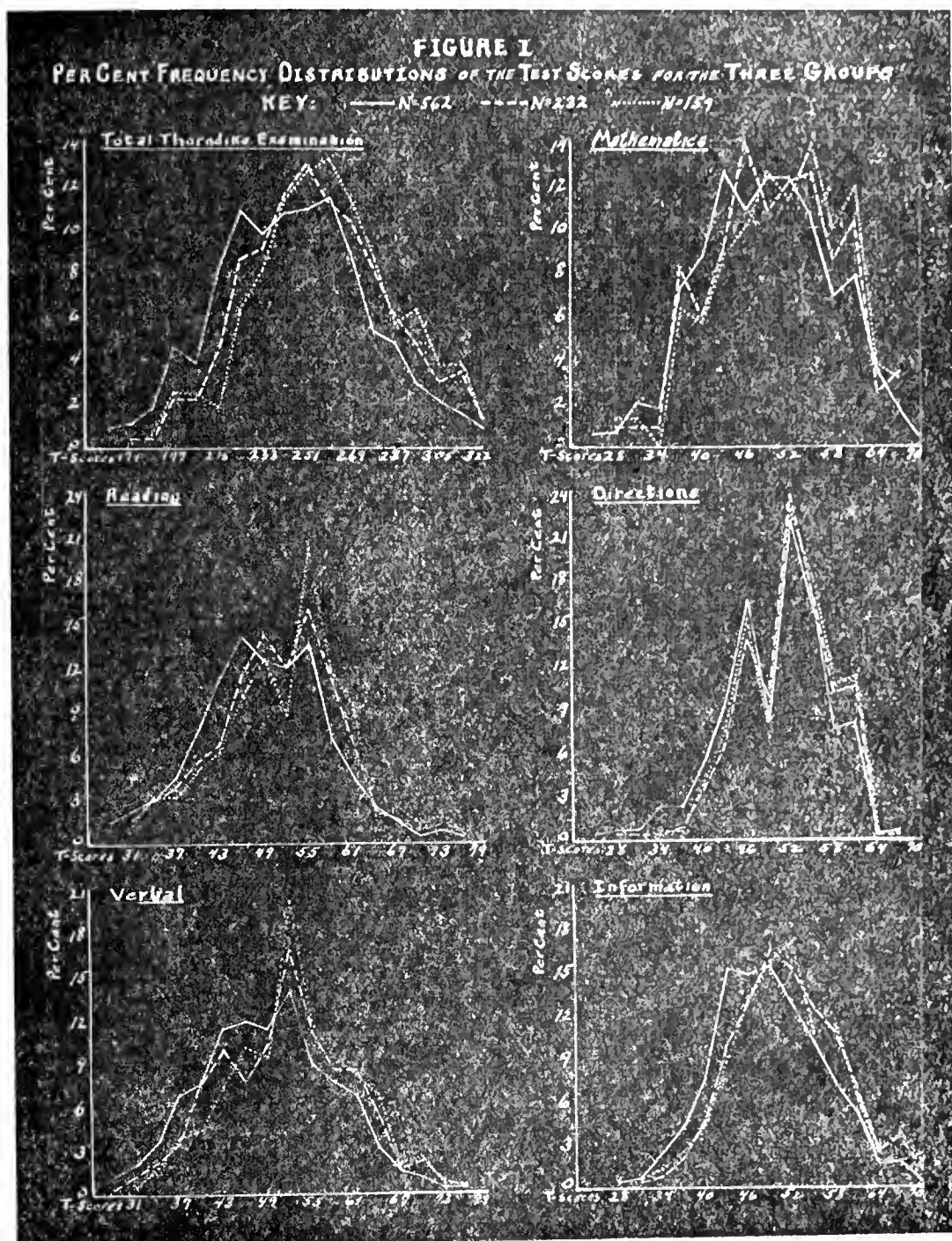
The frequency distributions of Figure I reveal the general similarity of the three groups. It is important to note that for the purposes of this comparison the 159 individuals are included in the two larger groups. The interest here lies in the comparison of the entering students, particularly those whose grade scores were used, with the total group that took the examination. For ease of comparison of the five test groups with each other, all of these distributions are in terms of per cent frequencies, based on the same group intervals, and derived from the reduced scores. The measures of skewness and their evaluation in terms of their standard error,<sup>6</sup> presented in Table IV, reveal only one case of decidedly significant (chances approximately 100 in 100) skewness. This case is the Following Directions tests,  $N = 562$ . One of the two tests comprising this group gave curve tending to bimodality, not only in the case of the 562 subjects but also for those entering the college. There was a marked tendency for these individuals either to do well or to do very poorly, on this test No. 8, Booklet I. It is one test that undoubtedly could be eliminated from the total examination; at least judging from the performance of these individuals such a step would be very practical.

Particularly satisfactory for the present analysis and general problem of this study is the fact that the chances are favorable for no *true* skewness for any of the tests or the

$$^6 \text{Sk} = P_{.50} - \frac{1}{2} (P_{.90} + P_{.10}) \qquad \sigma_{\text{sk}} = .59914 \frac{D}{N} \quad (17, \text{p. } 77)$$

Negative skewness indicates a piling of scores at the lower end of the distribution; positive skewness, at the upper end. For a symmetrical distribution,  $\text{Sk} = 0$ . A significant skewness is highly probable when the ratio of  $\text{Sk}$  to  $\sigma_{\text{sk}}$  is greater than three.

total examination in the cases of both the college groups. Also, in absolute amounts, the medians and percentiles of the college groups are not very different from those of the total



group. In view of the administrative use made of the examination there was the possibility of definite negatively skewed distributions. Actually, however, for the most reliable and significant groups of tests (Reading Comprehension, Verbal

TABLE IV  
MEDIAN, PERCENTILES (.90 and .10), SKEWNESS AND RATIOS OF SK TO  $\sigma_{SK}$   
FOR THE FIVE TEST GROUPS AND THE TOTAL THORNDIKE EXAMINATION

Test Variable	Median			P .90			P .10			Skewness			Sk/ $\sigma_{SK}$		
	562*	232*	159*	562	232	159	562	232	159	562	232	159	562	232	159
Reading Comprehension	49	51	53	60	61	61	38	39	41	0	1.0	2.0	0	1.2	2.1
Verbal Ability	49	52	52	61	63	63	37	40	40	0	0.5	0.5	0	0.6	0.5
Mathematical Ability	49	50	51	61	62	61	37	38	39	0	0	1.0	0	0	1.0
Directions	51	52	52	59	61	61	39	43	43	2.0	0	0	4.0	0	0
Information	48	51	51	59	62	62	39	42	42	-1.0	-1.0	-1.0	-2.0	-1.3	-1.1
Total Examination	246.3	254.5	257	287	297	298	206	220	223	-0.2	-4.0	-3.2	-0.1	-1.3	-0.9

\*562 = the total group taking the examination in June, 1925.  
232 = those entering Columbia College in the fall of 1925.  
159 = those whose three-year grade scores were available.

TABLE V  
MEANS, STANDARD DEVIATIONS, COEFFICIENTS OF VARIATION, RATIOS OF  $D/\sigma_D$ , AND RELATIVE VARIABILITY  
DIFFERENCES OF THE FIVE TEST GROUPS AND THE TOTAL THORNDIKE  
EXAMINATION FOR THE THREE GROUPS OF INDIVIDUALS

<i>Test Variable</i>	<i>Arithmetic Means</i>			<i>Standard Deviations</i>			<i>C. of Variation</i>			<i>Diff. <math>M_x - M_y / \sigma_D</math></i>			<i>Diff. <math>V_x - V_y</math></i>		
	<i>562(T)* 232(E)* 159(G)*</i>			<i>T E G</i>			<i>T E G</i>			<i>G-T E-T G-E</i>			<i>T-E T-G E-G</i>		
Reading	49.02	51.07	52.12	8.84	8.67	8.88	18.0	17.0	17.0	3.0	3.9	1.2	1.0	1.0	0.0
Verbal A.	48.85	51.24	52.13	9.19	9.14	8.89	18.8	17.8	17.1	3.3	4.1	1.0	1.0	1.7	0.7
Math. A.	48.82	50.16	50.95	8.75	8.58	8.60	17.9	17.1	16.9	2.0	2.8	0.9	0.8	1.0	0.2
Directions	50.06	51.87	52.02	7.40	6.36	6.13	14.8	12.3	11.8	3.5	3.4	0.2	-2.5	-3.0	-0.5
Information	48.84	51.54	51.63	7.86	7.77	8.14	11.8	15.1	15.8	4.4	3.8	0.1	-3.3	-4.0	-0.7
Total Exam.	245.59	255.72	258.84	30.83	29.55	29.19	12.6	11.6	11.3	4.3	5.0	0.8	1.0	1.3	0.3

\*T = the total group taking the examination in June, 1925. N = 562.  
E = those of the total group entering Columbia College in the fall of 1925. N = 232.  
G = those of the entering group whose three-year grade scores were available. N = 159.

Ability, and Number Ability) it was either zero or positive. For the total examination there is a tendency towards negative skewness.

The results of Table V reveal a great similarity in relative variability of the test scores for the three groups of subjects. In most cases there are only slight differences in the Coefficients of Variation of the college groups and the total group of 562 individuals. The greatest difference, General Information tests, between the 562 individuals and the 159 whose grades were used reveals relatively more spread for the scores of the 159 individuals than for the total group.

Although there are no reliable mean differences between the 232 freshman entrants and the 159 men whose three-year grades were used, there is evidently a difference reliably greater than zero between the means of the 159 and those of the 562. For the total examination scores, the chances of a difference between the means of these two groups are practically 100 in 100. However, such an estimate, although dependent on the Standard Deviation, is concerned with the difference between the typical score (mean) of each group. With ninety per cent of the total scores for the 562 candidates ranging from 194 to 302, and ninety per cent for the 159 three-year college grade group ranging from 209 to 305, and with a Coefficient of Variation in the former case of 30.83 per cent and in the latter of 29.19 per cent, it is very probable that the difference between the two groups is so small as to warrant the inference that the performance on the examination of the college groups are not, in respect to all of these considerations, very dissimilar to that of the total group of candidates taking the examination in June.

#### *B. Intercorrelations of Variables and Reliability Coefficients*

The Probable Errors of the intercorrelations coefficients of Table VI are of the order .02 or .03. The lowest coefficient, .1188, has a Probable Error of .028, barely one fourth as great as the coefficient. In fact all of the intercorrelations involving the fourth variable are low relatively and absolutely, when derived from raw scores. From the reduced scores, however, these intercorrelations become ten to twelve times greater than their Probable Errors, indicating that there is undoubtedly a reliable amount of correlation present when the scores on

TABLE VI  
INTERCORRELATIONS OF THE FIVE VARIABLES AND RELIABILITY  
COEFFICIENTS\*

Variables	1	2	3	4	5	Reliability C.	
						Raw	Reduced
1. Reading		.6673	.4603	.3681	.3968	.733	.708
2. Verbal A.	.6331		.4371	.3367	.4148	.765	.809
3. Math. A.	.4495	.4423		.3757	.3828	.557	.693
4. Directions	.2314	.1188	.1917		.2943	.408	.221
5. Information	.4009	.4040	.3389	.1507		.335	.455
Total Exam.						.759	.831

\*Coefficients derived from raw scores are below the diagonal, N = 568.

Coefficients derived from the reduced scores are above the diagonal, N = 562.

Coefficients used in the tetrad analysis are given to four places; others, only two or three places.

these tests are not subjected to constant errors possibly arising from differences in difficulty of the several forms of the examination booklets used. The theoretical reliability of this fourth variable is equal to .221, when derived from reduced scores. Any correlation between this variable and any other variable could not be greater than .4697 except by chance (.4697 being the square root of the reliability coefficient). The highest of the intercorrelations actually obtained with this fourth variable was equal to .3757.

The Mathematical Ability tests give an average correlation with the other four tests of .4140. Their correlation with the Verbal Ability tests is .4371 (reduced scores). This is in fairly marked contrast with Schneck's results (24, p. 23). The average intercorrelations he obtained were as follows: Verbal tests = .4920; Number tests = .3383; Verbal and Number tests = .1441. His further analysis confirmed the possibility suggested by these intercorrelations, viz., that his mathematical tests probably were measuring a factor common to all of them but not common to the verbal tests. The intercorrelations of this study suggest that the mathematical tests are measuring functions that overlap some of the functions measured by the other tests. Naturally, the more such overlapping, the less are these tests of an independent *mathematical ability*, assuming, of course, that the other four tests are not measures of an independent mathematical ability. In view of their content, such an assumption certainly seems very plausible.

The theoretical reliability of the total examination, deriving the coefficient from reduced scores, is very similar to that reported by Wood (35, pp. 45-51). Correlating the scores of the same group of students on two Forms of the Thorndike Examination, he obtained a reliability coefficient of .85; that obtained in this study is equal to .83. The Forms used by the students in the present study, it should be noted, represent the new Thorndike Examination entailing several changes over the old Form which Wood analyzed, and whose analysis was largely instrumental in bringing about the change. Consequently, the comparison of the reliability coefficient of the Thorndike Examination used in this study with that of the examination of Wood's study is not strictly of the same tests.

Using the same forms of the examination, and retesting after an interval of one year, Cowdery (4) obtained a reliability coefficient of .890. Using different forms, after an interval of one year, the coefficient was .751; for a two-year interval it was .720; and for a three-year interval it was .648. Cowdery advances the suggestion that this relatively low reliability coefficient (.648) might be attributable to the inclusion in the test of "much rather strictly scholastic material." Husband (14) comes to the defense of the reliability of the examination by attributing Cowdery's results to the relative homogeneity of the group taking the examinations. As a group becomes more homogeneous, slight individual differences tend to have a greater effect on the relative standing of individuals on a test. Consequently, according to Husband, the second Thorndike scores of these individuals with three years of varied experiences make the examination, "because of its very sensitiveness, appear to be unreliable."

The problem of the reliability of a reliability coefficient cannot exhaustively be dealt with here. The author, however, wishes to emphasize that the reliability coefficients obtained in this study were computed by the very frequently used method of correlating halves of a test and correcting for the number of cases by the Spearman-Brown "prophecy" formula (10, p. 269). This method is probably more reliable than that of retesting on the same test (17, p. 203), but it may not be as reliable as the method of retesting with a measure of "comparable difficulty." However, as Cowdery's results seem clearly to indicate, a reliability coefficient obtained by retesting is sub-

ject to difficulties and conditions that may make it as theoretical a coefficient as is the reliability coefficient obtained by the method of correlation of halves.

Concerning the reliabilities of the five groups of tests set up from the Thorndike Examination, it should be remembered that the author of this examination was concerned with the reliability of the total examination rather than with the reliability of various parts of it. The reliabilities of the Following Directions and General Information tests are considerably too low to give these tests much power to differentiate any abilities. In the case of the Mathematical Ability group the reliability coefficient derived from the reduced scores ( $r = .693$ ) is barely within the lower end of the range of coefficients indicative of satisfactory differentiating powers. Although reliability coefficients of .90 or above are most desirable, .80 is considered fairly practical, and .70 only probably practical. The reliability of the mathematical group of tests when determined from the raw coefficients becomes significantly less. This is apparently attributable to the fact that the reduced scores eliminated differences in difficulty for some of the mathematical tests of the three forms used.

The Verbal Ability tests, requiring only thirty-nine minutes of testing time, have a reliability coefficient, derived from reduced scores, that compares favorably with that of the total examination,—.809 as compared to .831. In fact, the Reading Comprehension tests plus the Verbal Ability tests have a reliability coefficient of .861 (Table IX). Although this coefficient is probably not *truly* higher than that for the total examination, it is *truly* as high.

The reliabilities of the first three variables are adequately high for a tetrad analysis based on these variables to have considerable significance. The reliabilities of the last two variables, however, are lower than is desirable. The problem here involves using these variables as they are, rather than devising highly reliable tests. Consequently, the groups of tests have been taken as found, and the results and interpretations characterized as estimates or in terms of their probabilities of truth, as they should be in any case.

### C. Tetrad Differences of the Five Variables

The tetrad differences obtained from the intercorrelation coefficients of the five groups of tests are presented in Table

VII. The Probable Error of the largest tetrad difference of each combination of the tests taken four at a time is also given. These Probable Errors were calculated from Kelley's formula, stated in Part II of this article.

TABLE VII  
TETRAD DIFFERENCES OF THE FIVE VARIABLES OF THE THORNDIKE EXAMINATION

Combination of Variables		From Raw Scores			From Reduced Scores		
		t <sub>abcd</sub>	t <sub>abde</sub>	t <sub>acdb</sub>	t <sub>abcd</sub>	t <sub>abde</sub>	t <sub>acdb</sub> *
1,2,3,4 <sup>c</sup>	tetrads =	.0680	.0191	— .0489	.0957	.0898	— .0059
	P.E. <sub>t</sub> =	± .0162			± .0168		
1,2,3,5	tetrads =	.0330	.0373	.0043	.0645	.0820	.0175
	P.E. <sub>t</sub> =		± .0158			± .0161	
1,2,4,5	tetrads =	.0019	.0478	.0459	.0473	.0628	.0191
	P.E. <sub>t</sub> =			± .0102		± .0166	
1,3,4,5	tetrads =	— .0107	— .0092	.0015	— .0054	— .0136	— .0082
	P.E. <sub>t</sub> =	± .0128				± .0139	
2,3,4,5	tetrads =	.0264	— .0108	— .0372	— .0003	— .0272	— .0269
	P.E. <sub>t</sub> =			± .0146		± .0141	

\*t<sub>abcd</sub> = r<sub>ab</sub>r<sub>cd</sub> — r<sub>ac</sub>r<sub>bd</sub>; t<sub>abde</sub> = r<sub>ab</sub>r<sub>ed</sub> — r<sub>ad</sub>r<sub>be</sub>; t<sub>acdb</sub> = r<sub>ac</sub>r<sub>bd</sub> — r<sub>ad</sub>r<sub>be</sub>  
<sup>c</sup>The tests designated by these numbers are named in Table VI.

In evaluating the probabilities that a given tetrad difference is not reliably greater than zero, Spearman usually judges that there is no significant difference, i.e., no difference reliably greater than zero, if the tetrad difference is less than five times its Probable Error. The present writer feels that an evaluation, more adequate for those not readily acquainted with the nature of the probabilities involved, would be in terms of the *actual chances* for a difference reliably greater than or equal to zero. He considers that procedure particularly important because one often loses sight of the fact that the number of chances in a hundred vary but slightly after the ratio of the difference and the Probable Error of the difference exceeds 3.00. For example, if this ratio equals 3.00, the chances are 97.9 in 100 of a true difference greater than zero. When the ratio is 4.00, they are 99.7 in 100, and when it is 5.00, they are 99.9+ in 100. Ordinarily, in statistical interpretation, a ratio of 4.00 is regarded as indicative of a satisfactory standard for a reliable difference. With regard to Spearman's criterion of

a ratio of 5.00, the writer does not take issue to the extent that he would maintain that a tetrad difference is *absolutely* greater than zero if it is barely less than five times its Probable Error. He does not maintain this because of the fact that there are still one or two chances left *in ten thousand* that no true difference exists. As a matter of probability, *absolute* certainty (exactly 100 chances in 100) is never reached unless the population or the measurements are infinite. All of which emphasizes the logarithmic relationship existing between the probabilities of a true difference and the ratio of the obtained difference to its Probable Error, and the consequent need of a working knowledge of the actual probabilities involved.

Inspecting the data of Table VII, it is evident that there is a decided tendency for the tetrad differences involving variables 1 and 2 (Reading Comprehension and Verbal Ability) to be reliably greater than zero, since the ratio of the differences to their Probable Errors is in most cases greater than 4.0,—thus leaving less than one half of one chance in 100 for no difference greater than zero. In fact, the only exception to this tendency is found in the case of the second combination of four tests ( $t_{1253}$ ), using the intercorrelation coefficients obtained from the raw scores. Here the ratio of .0373 to .0158 is 2.4, in which case there are five chances in 100 that the difference is insignificant. For this same combination of tests, using reduced scores, the ratio of the tetrad difference to its Probable Error is 5.1, there being about one-fiftieth of one chance in 100 that the difference is insignificant. In view of the nature of the reduced scores and the fact that these four tests had greater theoretical reliability coefficients when derived from the reduced scores than when derived from the raw scores, greater weight should undoubtedly be given to the tetrad differences obtained from the intercorrelations of the reduced scores.

Each set of three tetrad differences of the intercorrelations of the reduced scores involving variables 1 and 2 exhibit a relationship that tends to satisfy Kelley's XVI Proposition, stated in Part II of this article. In each of the three combinations of four variables involving variables 1 and 2, the first two tetrad differences are approximately equal to each other and reliably greater than zero, there being more than 99 chances in 100 for

such a difference. The third tetrad difference in each of these combinations is not significantly greater than zero, i.e., the chances are from nine to forty-nine in 100 that the true difference is zero,—assuming that their Probable Errors are of the order .02, a very safe assumption in view of the size of the Probable Errors obtained for the highest tetrad differences in each group. At least the probabilities are very great that they would not exceed .02 and not be less than .01.

When Kelley's XVI Proposition is satisfied there is evidence for the presence of a factor general to the four variables involved in the tetrad combination, *plus*, in addition thereto, a second factor common to variables 1 and 2 or variables 3 and 4, as well as the factors specific to each one of the four variables. In the three sets of tetrad differences under consideration, variables 1 and 2 are always represented by tests 1 and 2, i.e., the Reading Comprehension and the Verbal Ability tests. Variables 3 and 4 are represented by tests 3 and 4, and 3 and 5, or 4 and 5. It seems very probable, therefore, in view of the nature of the tests, that there is a factor here common to the Reading Comprehension and the Verbal Ability tests but not common to tests 3 and 4, 3 and 5, or 4 and 5. There is a possibility, however, that the group factor operating is common to tests 3 and 4, 3 and 5, or 4 and 5, since the satisfaction of Kelley's Proposition means that the group factor is common to variables 1 and 2 or to the other pair.

The most satisfactory analysis to determine in which pair the group factor is to be found cannot be made since at least six variables are necessary in order to take variables 1 and 2 with two variables other than 3 and 4, and 3 and 4 with two variables other than 1 and 2,—the procedure outlined by Kelley (18, p. 71). Presumptive evidence is obtainable, however, when five variables are available. Thus, using the five variables of this study, and making an analysis from the first four, i.e., variables 1, 2, 3, and 4, the group factor will lie in variables 1 and 2 if system A is satisfied; if system B is satisfied the group factor will lie in variables 3 and 4 (18, p. 73):

<i>System A</i>	<i>System B</i>
$t_{1234} = t_{1243} \neq 0$	$t_{1234} = t_{1243} \neq 0$
$t_{1235} = t_{1253} \neq 0$	$t_{1235} = t_{1253} = 0$
$t_{1245} = t_{1254} \neq 0$	$t_{1245} = t_{1254} = 0$
$t_{1354} = t_{1453} = 0$	$t_{1354} = t_{1453} \neq 0$
$t_{2354} = t_{2453} = 0$	$t_{2354} = t_{2453} \neq 0$

Evaluating the equality or lack of equality between the tetrad differences in terms of their Probable Errors, the relationship obtaining for the tetrads, from reduced scores, given in Table VII may most probably be characterized as follows:

$$\begin{aligned} (t_{1234} = .0957) &= (t_{1243} = .0898) \neq 0 \\ (t_{1235} = .0645) &= (t_{1253} = .0820) \neq 0 \\ (t_{1245} = .0473) &= (t_{1254} = .0628) \neq 0 \\ (t_{1354} = -.0136) &= (t_{1453} = -.0082) = 0 \\ (t_{2354} = -.0272) &= (t_{2453} = -.0269) = 0 \end{aligned}$$

It is evident that, evaluating the absolute sizes of the tetrads in terms of their Probable Errors, the results on the reduced scores fit system A. Consequently, there is presumptive evidence for a group factor common to the Reading Comprehension and Verbal Ability tests.

#### D. A Factor Common to the Thorndike Examination

If the test results of variables 1 and 2 are pooled and this combination taken as one variable and compared with the other three, theoretically the tetrad criterion should be satisfied for a common factor in case the group factor inferred in the preceding analysis actually exists through variables 1 and 2, since it would then become specific in the new set-up. Table VIII presents the results of such a procedure. For the intercorrelations of variable 1 plus variable 2 with the other variables, Spearman's method of correlating sums was used (25). As in Table VI, coefficients derived from the reduced scores are above the diagonal.

Taking the criterion that the *true* tetrad difference is equal to zero if the obtained difference is less than four times the Probable Error of the tetrad difference, it is evident that the intercorrelations of these four variables, which taken together make up the total Thorndike Examination, can be accounted for on the basis of a factor common to all four tests, plus specific factors for each test, and, of course, plus chance errors of measurement. This is the case for the intercorrelations from either the raw or the reduced scores. Making the estimation in terms of the actual chances that no tetrad difference is reliably greater than zero, for the largest ratio, 1.9, the chances are ten in 100 that the difference is insignificant. For the largest ratio, 1.5, of the tetrads derived from the reduced scores, the chances are sixteen in 100 that the difference is

TABLE VIII  
INTERCORRELATIONS, TETRAD DIFFERENCES AND THEIR PROBABLE  
ERRORS OBTAINED WHEN COMBINING THE READING  
COMPREHENSION AND VERBAL ABILITY TESTS

Variables	Intercorrelations				Reliability Coef.	
	I	II	III	IV	Raw Sc.	Reduced
I. Reading <i>plus</i> Verbal Ability		.4912	.3856	.4447	.845	.861
II. Math. Ability	.4910		.3757	.3828	.557	.693
III. Directions	.1831	.1917		.2943	.408	.221
IV. Information	.4444	.3389	.1507		.335	.455

Combination of Variables	From Raw Scores			From Reduced Scores		
	t	P.E. <sub>t</sub>	t/P.E. <sub>t</sub>	t	P.E. <sub>t</sub>	t/P.E. <sub>t</sub>
t <sub>I,II,III,IV</sub> =	.0119	±.0129	0.9	— .0030	±.0135	0.2
t <sub>I,II,IV,III</sub> =	— .0112	±.0152	0.7	— .0224	±.0146	1.5
t <sub>I,III,IV,II</sub> =	— .0231	±.0119	1.9	— .0194	±.0138	1.4

insignificant. The actual chances in 100 that the obtained tetrad is not truly greater than zero are thus probably high enough to justify the conclusion that the intercorrelations of these four groups of tests satisfy the criterion for a common general-factor plus specific factors.

Since these four groups of tests making up the Thorndike Examination satisfy the tetrad criterion, the problem now is to determine the extent to which these four groups of tests measure the common factor. With regard to the question, “Are the intellectual functions or abilities measured by the Thorndike Examination of such nature that they may be characterized as *general scholastic ability*?”, the results of this section indicate that these functions are so organized that a general factor may be thought of as running through the four groups of tests assembled. The question of the extent or size of this general factor will be dealt with in the next section. The further question of the meaning of this general factor, whether or not it may be identified with *general scholastic ability* will be considered in Sections G and H.<sup>7</sup>

<sup>7</sup> An analysis of tetrad differences derived from correlation coefficients corrected for attenuation is not made in this study for two reasons: (1) The formula for the Probable Error of a corrected tetrad is not available in a usable form; without the Probable Errors of the corrected tetrads they cannot be properly evaluated; and (2) Were the formula available, it would hardly be advisable to use corrected coefficients here because of the low reliabilities of two of the variables. The tetrad differences from corrected correlation coefficients, representing the intercorrelations of the four variables analyzed in this section, were calculated and found to be as follows:

Raw scores:      t'<sub>1234</sub> = .0470;      t'<sub>1243</sub> = —.0442;      t'<sub>1342</sub> = —.0913  
Reduced scores:    t'<sub>1234</sub> = —.0123;      t'<sub>1243</sub> = —.0911;      t'<sub>1342</sub> = —.0793

### E. *Relative Importance of the Common and Specific Factors*

In attempting to determine the relative roles of the common factor and the factors specific to each of the four groups of tests that have been assembled from the total Thorndike Examination, the problem has been attacked as follows:

1. What per cent of the variance of each of the four tests is attributable to the common factor, which will be called  $c$ ? This can be *estimated* from the coefficient of correlation between each test and  $c$ ,  $r_{xc}$ . The per cent of variance determination not attributable to  $c$  will be attributable to  $s + e$ , where  $s$  represents the specific factors and  $e$  the chance errors of measurement (the variance here will be partially determined by chance errors since these coefficients have not been corrected for attenuation). The coefficient  $r_{xc}$  is obtained, when there are four variables, by the following formula derived by Spearman (26, p. xvi, Formula 20):

$$r_{xc}^2 = \frac{r_{xw}r_{xy} + r_{xw}r_{xz} + r_{xy}r_{xz}}{r_{wy} + r_{wz} + r_{yz}}$$

In this formula,  $x$  represents the variable or test under consideration;  $c$ , the common factor, is represented by  $g$  in Spearman's formula<sup>8</sup>;  $w$ ,  $y$ , and  $z$  represent the three other variables of the tetrad. Root  $r_{xc}^2$  represents the correlation obtaining between the variable and the common factor. As is evident, this formula rests on the assumption that the correlation between the specific factors of two variables is equal to the partial correlation between the two variables, with  $c$  held constant. This partial correlation is taken to be zero on the assumption that the specific factors do not correlate with each other. For this to be true, the tetrads have to equal zero, or to be probably true they have to be not reliably greater than zero in terms of their Probable Errors.

The correlations obtained for each variable with  $c$  ( $r_{xc}$ ) are presented in Table IX, as also are the intercorrelations of the variables with  $c$  partialled out.

The correlations with the common factor  $c$  are about the same, whether the raw scores or reduced scores are taken, except in the case of variable III. Since this one is .5290, when

<sup>8</sup> The present writer does not wish this common factor  $c$  to be confused with Spearman's  $g$ , since it is very improbable that they might have identical meanings.

TABLE IX  
CORRELATIONS WITH C AND INTERCORRELATIONS  
WITH C PARTIALLED OUT\*

Variable	From Raw Scores				From Reduced Scores			
	$r_{xc}$	Partial Coefficients			$r_{xc}$	Partial Coefficients		
		II	III	IV		II	III	IV
I°	.7557	.0014	— .0252	.0552	.7418	— .0231	— .0120	.0333
II	.6472		.0184	— .0486	.6776		.0276	— .0111
III	.2680			.0003	.5290			— .0141
IV	.5576				.5748			

\* $r_{xy.c} = r_{xy} - r_{xc}r_{yc}/k_{xc}k_{yc}$ ,  $k$  being the coefficient of alienation.  
°The tests designated by these numbers are named in Table VIII.

derived from reduced scores, it is very probable that the four correlations based on reduced scores best support the contention, made in the previous section, that there is a factor common to all of these tests. If the intercorrelations of the variables with  $c$  partialled out are not reliably greater than zero, they support the contention that there are no group factors operating through these tests. In absolute size, all are practically zero, and it is very probable that their *true* value is exactly zero. The exact probabilities are not given since a formula for the Probable Error of  $r_{xc}$  is not available.

These results and those of the tetrad analysis certainly suggest that there are no group factors operating significantly. Such an assumption being made, the problem, as stated at the beginning of this section, is to determine the per cent of the variance of each of the four tests attributable to the common factor,  $c$ . The *estimate* of the per cent of various attributable to  $c$  is obtainable from  $r^2_{xc}$ , since  $r^2_{xc} = r_{\sigma_x\sigma_c}$  (17, p. 178, Formula 121; 7, p. 375, Note 4). This formula is derived on the assumption of rectilinearity, homoscedasticity, and equal kurtosis of the correlation surface. It cannot be asserted that these conditions hold absolutely here, nor can their exact probability of holding be determined. The assumption, therefore, will be made that these conditions hold adequately for the purposes of this analysis. The estimates of the per cent of variance as given by  $r^2_{xc}$  are presented in Table X.

That the per cent of variance attributable to  $s + e$  is equal to 100 per cent minus the per cent attributable to  $c$  is an artifact of the situation, since the per cent attributable to  $s + e$  is taken to be the coefficient of alienation squared,  $k^2_{xc}$  being

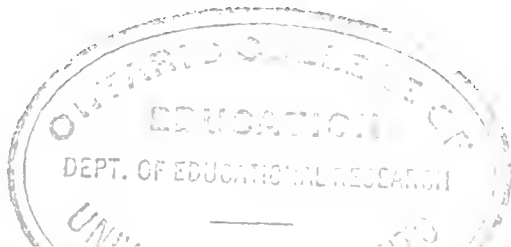
TABLE X  
ESTIMATED PER CENT OF VARIANCE OF THE VARIABLES  
ATTRIBUTABLE TO  $C$  AND  $S + E$

Variables	From Raw Scores		From Reduced Scores	
	% to $C$	% to $S + E$	% to $C$	% to $S + E$
Reading plus Verbal A.	57.1	42.9	55.0	45.0
Mathematical Ability	41.9	58.1	45.9	54.1
Following Directions	7.2	92.8	28.0	72.0
General Information	31.1	68.9	33.0	67.0

equal to  $1.00 - r_{xc}^2$ . Spearman gives  $r_{xs}^2$  as equal to  $1.00 - r_{xc}^2$  (i.e.,  $k_{xc}^2$ ) on the assumption that  $r_{xs \cdot c} = 1.00$ , and  $r_{sc} = 0$ . This correlation of  $r_{xs \cdot c}$  will hold when the tetrads are equal to zero, provided none of the variance of  $x$  is determined by chance errors. Since chance errors do determine some of the variance of  $x$  in the present analysis, the correlation may be characterized more truly as follows:  $r_{x[s + e] \cdot c} = 1.00$ . The assumptions underlying the use of  $k_{xc}^2$  as representing the per cent of variance of a variable attributable to  $s + e$  are the same as those made in using  $r_{xc}^2$  as representing the per cent of variance of a variable attributable to  $c$ .

The results presented in Table X indicate that the common factor plays a slightly more important part in determining the variance of variable I than do the specific and chance factors. Relatively,  $c$  also plays a fairly important part in the determination of the variance of the other three variables, particularly in the case of the reduced scores. In these three tests, however, the specific and chance factors have a greater weight of determination. What, then, is the relative importance of  $s$  in determining the variance of the variables? An answer to this question is made in the following manner:

If the tetrad differences were truly equal to zero, and if  $x$  is taken to represent one of the variables, then  $x = kc + s + e$ , where  $k$  is a constant that determines the relative proportion of  $c$  to this variable. On the assumption that the tetrads of this study were truly equal to zero, if variable I of Table X is taken to represent  $x$ , the variance of  $I = r_{ic}^2 + k_{ic}^2$  ( $k^2$  being the coefficient of alienation squared). Assuming that the variance determinations of  $c$  actually obtained are truly applicable, then the variance of  $I = 55.0\% + 45.0\%$ . In which case 55.0 per cent of the variance of variable I is attributable



to  $c$  and 45.0 per cent to  $s + e$ . Now, if the total variance of variable I is attributable to the operation of true factors plus the operation of chance factors, the square root of the reliability coefficient *squared* will represent the per cent of the variance attributable to true factors and  $1 - \sqrt{r_{11}}$  <sup>2</sup> (numerically, of course this is the same as  $1 - r_{11}$ ) will represent the per cent of the variance attributable to chance factors (18, p. 36; 10, pp. 272-273). Thus by using the theoretical reliability coefficients obtained for the four variables under consideration here (see Table VIII), Table XI can be set up to give *estimates* of the per cent of the variance of these variables attributable to  $c + s$ , and the per cent attributable to  $s$  alone, since estimates of the per cent attributable to  $c$  have already been made. The per cent attributable to  $e$  will, of course, be the difference between that attributable to  $c + s$  and 100 per cent.

TABLE XI  
ESTIMATED PER CENT OF VARIANCE OF THE VARIABLES  
ATTRIBUTABLE TO  $C + S$ ,  $S$  AND  $E$

Variables	From Raw Scores			From Reduced Scores		
	$C + S$	$E$	$S^*$	$C + S$	$E$	$S^*$
Reading plus Verbal A.	84.5	15.5	27.4	86.1	13.9	31.1
Mathematical Ability	55.7	44.3	13.8	69.3	30.7	23.4
Following Directions	40.8	59.2	33.6	22.1	77.9	0
General Information	33.5	66.5	2.4	45.5	54.5	12.5

\*Per cent attributable to  $s$  is taken as equal to  $(C + S) - C$ .  
(See Tables VIII and X.)

As indicated in the analysis leading up to the results of Table XI, these estimates of the per cent of variance of each test attributable to the common and specific factors are truly valid only to the extent that certain conditions are fulfilled. Were the reliabilities of tests III and IV .70 or higher, it could be affirmed with a greater degree of assurance that these estimates approximate fairly closely the *true* relationships. The suggestion, however, seems to be clear that the common factor functions relatively greater than the specific factors in determining the variance of each variable, and that the per cent of variance of variables III and IV attributable to chance factors is very great.

2. What per cent of the variance of  $c$ , the common factor,

is attributable to each variable and to the whole Thorndike Examination? These per cents of determination can be estimated from the Beta coefficients of the regression equation, in which each variable is expressed in terms of its own standard deviation. The product of a partial Beta coefficient and the correlation coefficient of the particular independent variable and dependent variable represents the per cent of variance of the dependent variable (in this case,  $c$ ) attributable to the independent variable. The sum of these products is equal to the squared multiple  $R$ , viz.,  $R^2_{c \cdot I, II, III, IV}$ . Table XII presents the results obtained by this procedure.

TABLE XII

ESTIMATED PER CENT OF VARIANCE OF  $C$  ATTRIBUTABLE TO EACH VARIABLE AND TO THE THORNDIKE EXAMINATION AS A WHOLE

Variables	From Raw Scores		From Reduced Scores	
	Beta* Coefficient	Per Cent of Variance	Beta* Coefficient	Per Cent of Variance
Reading plus Verbal Ability	.4491	33.24	.4187	31.06
Mathematical Ability	.3371	22.32	.3218	21.81
Following Directions	.1700	4.58	.1845	9.76
General Information	.2129	11.90	.2111	12.14
Total Examination		72.04		74.77
$R^2_{c \cdot I, II, III, IV} =$		.7204		.7477
$R_{c \cdot I, II, III, IV} =$		.8488		.8647

\*Instead of deriving these Beta coefficients from the  $b$  coefficients of the regression equations, they were obtained directly from the solutions of the simultaneous equations set up in terms of correlation coefficients rather than product moments. The equations were solved in the manner outlined by Ezekiel (7, pp. 165-9).

According to the estimates of Table XII, nearly half of the variance of the common factor attributable to the whole examination may be assigned to variable I, i.e., the Reading Comprehension and Verbal Ability tests, and more than two-thirds of the variance to variable I plus variable II (Mathematical Ability tests). Perhaps the most significant estimate is that of the determination of the total examination, according to which (from reduced scores) seventy-five per cent of the variance of the common factor is attributable to the total Thorndike Examination. Since the reliability of the total examination was found to be equal to .83, and the square root of this reliability coefficient *squared* is equal to the per cent

of variance of the total examination attributable to *c* and *s* factors, the estimate may be approximated that the total examination is functioning in the following manner:

(a) 75 per cent of its variance is attributable to a common factor.

(b) 10 per cent of its variance is attributable to specific factors.

(c) 15 per cent of its variance is attributable to chance errors.

Although these estimates are necessarily qualified by the assumptions underlying them, they may be taken as very suggestive of the manner in which the Thorndike Examination is functioning, since there is a fair probability that some of the important conditions are adequately satisfied, e.g., the tetrad differences were revealed as having fair probabilities of equalling zero. Were the reliability coefficients *true* measures, rectilinearity, homoscedasticity, and equal kurtosis *truly* characteristic of the interrelated distributions, and the tetrad differences *truly* zero, the above determinations would be most probably the *true* estimates.

That only a very small per cent of the variance of the total examination is probably determined by a specific factor is supported by the principle that a total score on an examination testing general and specific functions tends to maximize the general functions and minimize the specific.

## F. *Relationship of the Test Variables to Scholastic Records*

### 1. *Adequacy of the Sample:*

The results and interpretation thereof in support of the contention that of the total group taking the Thorndike Examination in June, 1925, those entering Columbia College in the fall of that year were very similar to the total group in respect to their performance on the total examination and on the five groups of tests assembled from the examination are presented in Section I of *The Results*. The present writer wishes again to emphasize that this comparison was made in order to support the reasonableness of any inferences drawn from the following analysis that may throw some light on the nature of the factor inferred to be common to the examination and fairly extensively measured by it.

All of the correlations with the various grade criteria are made on a population of 159 boys, each of whom had six hours or more work in each of the five grade division, during his first three college years. The Contemporary Civilization grades represent at least ten hours work, taken during the freshman year.

2. *Intercorrelation and Reliability Coefficients of the Test Variables for the College Group:*

Table XIII presents the intercorrelation coefficients of the groups of tests assembled from the Thorndike Examination and the reliability coefficients of these groups of tests, as derived from the 159 cases.

TABLE XIII  
INTERCORRELATION AND RELIABILITY COEFFICIENTS OF THE  
THORNDIKE EXAMINATION SCORES FOR THE COLLEGE GROUP\*

Variables	Intercorrelation Coefficients (N=159)						Reliability Coefficients	
	1	2	I	II	III	IV	Raw	Reduced
1. Reading C.		.61		.37	.30	.39	.719	.689
2. Verbal Ability	.57			.33	.23	.39	.726	.770
I. Reading + Verbal				.37	.28	.42	.821	.844
II. Math. Ability	.37	.35	.40		.44	.38	.509	.654
III. Directions	.27	.03	.15	.16		.19	.361	.056
IV. Information	.37	.43	.45	.34	.03		.394	.431

\*Coefficients above the diagonal are from reduced scores.

These intercorrelation coefficients are very comparable to those obtained from the total group taking the examination. In most cases they are slightly lower, and the largest differences are equal to about 2 P.E. of their difference.

The reliability coefficients also are very comparable. They are slightly lower for the college group than for the total group of candidates, all differences, however, being less than 1 P.E. of the difference, with one exception, viz., the reliability of variable III for the college group is practically zero, whereas it is equal to .221 for the original group. This is the variable in which one of the tests was found in the original group to be quite worthless from practically all points of view; its worthlessness is accentuated by this result with the college group.

On the whole, the intercorrelations of the variables for the college group, being very similar to those of the total group,

give further support to the contention that the performance of the two groups is very similar. This contention is further attested to by the reliability coefficients of the total examination, which were found to be, from reduced scores, exactly the same, i.e., to four places, for the original group and for this college group. They were in both cases equal to .831. That they are exactly the same is a coincidence, but that they are practically the same is undoubtedly largely attributable to the similarity of the performance of the two groups.

### 3. *Intercorrelation Coefficients, Means, and Standard Deviations of the Grade Scores:*

Table XIV presents the intercorrelations, means, and standard deviations of the scores for the various grade criteria derived from three-years' scholastic records of the college group of 159 individuals.

TABLE XIV  
INTERCORRELATION COEFFICIENTS, MEANS, AND S. D.'s  
OF THE GRADE SCORES

<i>Variable</i>	<i>S.</i>	<i>C.C.</i>	<i>S.S.</i>	<i>E.</i>	<i>F.L.</i>	<i>Mean</i>	<i>S.D.</i>	<i>C.V.*</i>
Total Grades	.871	.776	.816	.697	.798	5.72	1.71	30.0
Science & Math.		.612	.683	.535	.642	5.54	2.43	43.9
Contemp. Civiliz.			.652	.565	.561	5.94	2.17	36.5
Social Science				.583	.582	5.84	1.86	31.8
English					.557	5.58	1.46	26.2
Foreign Language						5.80	2.12	36.5

\*Pearson's Coefficient of Variation.

The theoretical reliability of the total grade scores can be estimated from the intercorrelations of grades in Science and Mathematics, Social Science, English, and Foreign Language, since each division is based on a three-year grade average for each individual of the group of 159. The grades of these four variables represent for practically all of the subjects over eighty per cent of their total grade score for the three years. Taking the Science and Mathematics plus English grades with Social Science and Foreign Language grades, the correlation is .7962. Estimating from this by the Spearman-Brown prophecy formula, the theoretical reliability coefficient of the grades is .8866.

A coefficient as high as this is not quite in line with the often quoted statement to the effect that grades are notoriously un-

reliable. Usually the reliability of grade scores is estimated by correlating one semester's grades with those of another semester. Crawford (5), for example, reports a correlation of first and second term grades equal to .85. The manner of determination used in this study, however, has the advantage of taking into the estimate the time factor as a continuous function. Furthermore, in spite of the different standards of grading used by different departments, it seems here that the relative standing of the individuals in the several departments is fairly similar. On the other hand, over a three-year period, constant errors of teachers' judgments enter into the determination; for example, certain individuals acquire the reputation of doing poorly, average, or very well. This disadvantage, however, is not eliminated by correlating one semester's grades with those of another semester, nor does this method have the two advantages accruing to the other method. It is very probable, therefore, that the method of estimation used in this study is the *more* reliable way of determining the reliability coefficient.

#### 4. *Correlations between the Examination Scores and Grades:*<sup>9</sup>

Table XV presents the correlations between various parts and combinations of parts of the Thorndike Examination and grades and the correlations of the total examination with grades.

The population from which the correlation coefficients of Table XV were derived being 159, a coefficient to be four times

---

<sup>9</sup> A survey of the literature on the relationship between psychological examinations and grades reveals correlations that range usually from about .25 to .60. Some of the articles giving general summaries of many of these studies are by Thurstone (31), Pintner (23), Whipple (33), and MacPhail (20). Thurstone, for example, reports results on 5,200 students in 26 institutions giving American Council tests; the correlations reported ranged from .23 to .57, averaging .45.

Grauer and Root (12) report a study on the relationship of the Thorndike Examination (old form) and grades: for 569 freshmen, the correlation with first semester grades was .51; for 159 subjects, with first two-year grades, it was .39. After an extensive analysis of the predictive value of the examination, they conclude that "the correlation between the Thorndike score and the average academic grade is too low to justify the exclusion of students from college on the basis of the Thorndike rating alone."

Wood (35) reports correlations (old form of the Thorndike Examination) as high as .67, obtaining this coefficient when discarding the records of all men late to the examination, of those not native to the English language, of those of long illnesses, of certain disciplinary cases, of voluntary withdrawals, and of all admitted under the Old Plan. He

its Probable Error has to be larger than .20. All of the Probable Errors are of the order .05 or .04, e.g.,  $r = .07 \pm .05$ ,  $r = .40 \pm .045$ ,  $r = .50 \pm .04$ .

The results of Table XV may be summarized as follows:

1. All of the correlations of grades with the Reading Comprehension tests, with the Verbal Ability tests, or with the combination of both are greater than four times their Probable Error,—most of them are six to ten times greater.

2. All of the correlations of grades with the Mathematical Ability tests are four to seven times greater than 1 P.E., with the exception of the correlations of Foreign Language grades.

3. All of the correlations of grades with the Following Directions tests are less than 4 P.E., with four exceptions. The largest of these four is .31, with Science and Mathematics grades.

4. All of the correlations of grades with the General Information tests are less than 4 P.E., with the exceptions of the correlations with Contemporary Civilization grades.

5. Comparing the correlations derived from raw scores and from reduced scores, there are no cases in which the difference between pairs is greater than four times the Probable Error

---

concludes that "the intelligence test is not only as good a criterion for admission to college as any other single criterion thus far used (written in 1923), but it is more efficient and less expensive."

Lauer and Evans (19) report a correlation of .42 for intelligence test scores and first quarter grades, a correlation of .49 for High School averages and first quarter grades, and a multiple correlation of these two criteria with grades of .55. Guiler (13) reports correlations ranging from .40 to .52. Cleeton (3) obtained correlations between Thorndike scores (old form) and grades for various groups of subjects, ranging from .38 to .52; correlations between the Iowa High School Content Examination and grades of around .50; and a correlation of .61 between grades and Thorndike plus the Content Examination scores.

Edgerton (6) reports correlations between three-year scholarship and the Ohio University Intelligence Examination ranging from .40 to .64. The average correlation was about .50. Correlations between first quarter grades and three-year scholarship averaged about .88. The first quarter's grades, however, were included in the three-year scholarship grade. The multiple correlation of first quarter grades *plus* the intelligence test scores with three-year scholarship also averaged about .88. The relatively high predictability of three-year scholarship from first quarter grades and intelligence test scores is emphasized by Edgerton. It is observed, however, that the predictability of three-year scholarship from first quarter grades *alone* is practically as great as from first quarter grades plus intelligence test scores. Obviously, the exceedingly greater predictability of three-year scholarship from first quarter grades than from the intelligence test scores can be taken advantage of only for those students *admitted* to the college and completing the first quarter's work.

TABLE XV  
CORRELATIONS OF THE EXAMINATION SCORES WITH GRADE SCORES  
N = 159

Thorndike Examination Variables	Total Grades		Sci. & Math.		Con. Civliz.		Social Sci.		English		Foreign Lang.	
	Raw	Redc.	Raw	Redc.	Raw	Redc.	Raw	Redc.	Raw	Redc.	Raw	Redc.*
1. —Reading	.33	.34	.21	.23	.45	.45	.33	.33	.35	.34	.22	.23
2. Verbal A.	.35	.36	.23	.24	.39	.41	.35	.36	.38	.39	.31	.28
I Reading+Verbal	.39	.37	.25	.25	.47	.46	.39	.36	.41	.38	.31	.27
II Math. A.	.32	.30	.34	.33	.27	.26	.28	.26	.23	.22	.18	.17
III Directions	.20	.29	.22	.31	.11	.15	.13	.23	.07	.14	.08	.16
IV Information	.20	.13	.13	.10	.30	.22	.19	.10	.14	.09	.20	.09
I+II	.42	.41	.34	.32	.46	.46	.40	.39	.40	.39	.31	.28
I+II+III	.44	.43	.37	.36	.45	.44	.41	.40	.39	.37	.30	.28
I+II+III+IV	.41	.37	.33	.29	.46	.45	.38	.35	.35	.34	.31	.26
1.+1.+2.+II+III+IV	.42	.40	.32	.32	.48	.46	.39	.43	.37	.36	.30	.27
1/2(I+II+III+IV)	.40	.43	.29	.34	.47	.46	.39	.40	.35	.36	.33	.31
Thorndike Score	.41		.31		.46		.37		.36		.30	

\*"Raw" refers to the correlations with the raw scores of the examination.  
"Redc." refers to the correlations with the reduced scores of the examination.  
1+II+III+IV represents the sum of the scores for the total examination.  
1.+1.+2.+II+III+IV represents the sum of the scores for the total examination with the Reading Comprehension scores taken twice.  
These latter scores are given this weight in the regular Thorndike Score.  
Thorndike Score represents the regular Thorndike Score as recorded in the Admissions Office. It is a transmuted total examination score.

of the difference. In fact, there are very few cases in which the difference is greater than 1 P.E. the difference. Practically all of the exceptions to the latter statement are found in the correlations involving variables III or IV.

6. The correlations of grades with variable I (Reading Comprehension plus Verbal Ability) are practically of the same size as those of grades with the total examination, regardless of which of the five criteria of the total examination score is considered. In fact, none of the differences is greater than 1 P.E. its difference, except in the case of the correlations with Science and Mathematics grades, where, however, none of the differences is greater than 2 P.E. the difference.

7. None of the correlations of grades with either part of variable I, i.e., the Reading Comprehension tests or the Verbal Ability tests, is greatly different from the correlations of grades with the total examination. In fact, none of the differences is greater than 2 P.E. the difference, except in the case of the correlations with Science and Mathematics grades, where none of the differences is greater than 3 P.E. the difference.

8. The correlations of the Mathematical Ability tests with Science and Mathematics grades are practically the same as the correlations of the total examination with Science and Mathematics grades. Any differences are less than 1 P.E. the difference. None of the correlations of other grade criteria with the Mathematical Ability tests is very similar to the correlations of these grade criteria with the total examination. For example, Mathematical Ability and Contemporary Civilization grades,  $r = .27$ ; total examination and Contemporary Civilization grades,  $r = .45 - .48$ , being the range of the six coefficients. The difference is about 3 P.E.

9. In several instances the correlations of grades with variables I + II are higher than the correlations of grades with the total examination. However, the differences are less than 1 P.E. the difference.

10. The correlations of one half of the total examination, being the sum of one half of the tests comprising each variable (i.e., four of the Reading Comprehension tests, three of the Verbal Ability tests, etc.) with the various grade criteria are practically the same as the correlations of the total examina-

tion with grades. All differences are less than 1 P.E. the difference and most are less than .5 P.E. the difference.

11. The correlations of the total examination with the various grade criteria are practically the same for all five total examination scores, i.e., raw scores and reduced scores with the Reading Comprehension test scores weighted twice, and unweighted, and the Thorndike transmuted scores. All differences are less than 1.5 P.E. of the difference; practically all are less than 1 P.E. the difference.

From this summary of the correlations of Table XV, the following inferences are made:

*To the Extent That a Correlation Coefficient in Itself Is an Index of Predictability,*

1. The various grade criteria, with the exception of Science and Mathematics grades, can be predicted as well from variable I (Reading Comprehension plus Verbal Ability tests) as from the total examination. This inference is also supported by the theoretical reliability coefficients: from reduced scores, the reliability coefficient of variable I was .844, of the total examination it was .831. Thus, the reliability of variable I is as great as that of the total examination.

2. The various grade criteria, with the exception of Science and Mathematics grades, can be predicted nearly as well from *either* the Reading Comprehension tests *or* the Verbal Ability tests as from the total examination. The reliability coefficients of each of these halves of variable I are not as large, however, as the reliability coefficient of the total examination,— .689 and .770 as to .831.

3. The Following Directions tests not only have no predictive significance but their inclusion in the total examination lowers its reliability slightly.

4. The General Information tests have no predictive significance. The only correlation of this variable revealing a reliable amount of relationship is that with Contemporary Civilization grades, indicating probably that part of this variable is functioning specifically. However, it is necessarily a relatively small part.

5. *Half* of the Thorndike Examination has as high a value of prediction for any of the various grade criteria as has the

entire examination. That the sum of half of each of the five variables would serve all of the functions of the total examination is probable, but not necessarily true. Such a possibility is certainly suggested by these results, but, for one thing, the reliability of half of the examination is undetermined.

#### G. *The Probable Nature of the Common Factor. I*

What light do the results of the two preceding sections throw upon the nature of the factor common to the total examination? In view of the fact that the correlations of variable I (Reading Comprehension plus Verbal Ability tests) with total grades and with the other grade variables, except the Science and Mathematics variable, are practically the same as those of the total examination with total grades and these other grade variables; and in view of the fact that nearly half of the variance of the common factor assigned to the total examination was estimated to be attributable to variable I, it would seem that the examination is measuring intellectual functions that might best be characterized as verbal ability *plus* certain factors dependent upon the testing situation.

The estimate was made that seventy-five per cent of the variance of the total examination is attributable to a common factor. This common factor, then, can probably be characterized in terms of a knowledge of words and verbal relationships and the ability to use them *plus* factors common to the total testing situation, such as habits of speed in a competitive situation calling for the manipulation of much verbal material, pencil-paper tasks, ability to adapt to a highly motivated group contest, and general environmental conditions. That these factors might be included in *the common factor* is very probable since they are factors that could operate continuously and relatively the same to the individuals during the whole examination period. Furthermore, of these several factors, the one probably most determining the correlation found with the various grade criteria is the *verbal ability* factor, since it would be a factor necessarily common to all of the grade scores.

A further estimate was made, viz., that ten per cent of the variance of the total examination is attributable to a specific factor. In view of the differential relationship of the Mathematical Ability tests with Science and Mathematics grades, it is probable that a part of this specific factor for the total ex-

amination, although a relatively small part, can be characterized in terms of the ability to manipulate numbers and number concepts. However, in view also of the intercorrelations of the Mathematical Ability tests with the other test variables, and in view of the correlations with other grades, and since more of the variance was attributable to the common factor than to the specific factor, it is very probable that this *group* of Mathematical Ability tests is measuring more of the common factor than number ability.

Since the grade criteria and the total Thorndike Examination correlated from only .40 to .50, and since the degree of concomitance may be largely attributed to the operation of a verbal ability factor, the grade scores are largely determined by factors other than those of verbal ability, *as it is measured by the examination*. Aside from chance and constant errors entering into the judgments of those making the grades, other factors that probably have relatively important functions in the determination of the grades are factors dependent upon curricular activities, extra-school work, health, motivation differences, freedom from economic and other worries, and moral differences [Cf. Freeman's article (9) and May's (21)]. But, in addition, other factors that probably have relatively important functions in the determination of the grades are factors not only dependent directly upon the particular subject matters of the various studies but also cognitive *group factors*, other than verbal ability, such as number ability (24) and memory ability (1).

#### H. *The Probable Nature of the Common Factor. II*

A more reliable judgment of the nature of the common factor than that made on the basis of the correlation coefficients of the test and grade variables can be had from the estimates of the per cent of variance of the various grade criteria attributable to each test variable as well as to the total examination, giving each test its best weight. These estimates can be made from the partial Beta coefficients and the multiple correlation coefficients of the examination with the various grade criteria.

Table XVI presents the Beta coefficients and multiple correlation coefficients obtained. Both the Beta coefficients and multiple correlations were derived in the manner used to obtain the Beta coefficients and multiple correlations of Table XII. Table XVII presents the estimates of the per cent of the vari-

ance of each grade variable attributable to each test variable as well as to the total examination.

TABLE XVI  
PARTIAL BETA COEFFICIENTS AND MULTIPLE CORRELATIONS  
OF THE TEST VARIABLES WITH THE GRADE CRITERIA

<i>Variable</i>	<i>Scores</i>	<i>Total G.</i>	<i>Con. Civ.</i>	<i>Sci. Math.</i>	<i>English</i>	<i>Soc. Sci.</i>	<i>For. Lang.</i>
1. Reading C.	Raw	.1334	.3463	— .0037	.2171	.1779	.0615
	Redc.	.1481	.3055	.0419	.1684	.1414	.0778
2. Verbal A.	Raw	.2507	.1632	.1732	.3073	.2510	.2506
	Redc.	.2109	.2003	.1361	.2999	.2611	.2178
II. Math. A.	Raw	.2252	.1277	.3317	.1268	.1855	.0826
	Redc.	.1493	.1029	.2182	.1172	.1330	.0512
III. Directions	Raw	.1811	.0200	.1980	— .0341	.0889	.1366
	Redc.	.1499	— .0351	.1877	— .0086	.0931	.0798
IV. Information	Raw	— .0934	.0441	— .0664	— .1406	— .1313	— .0492
	Redc.	— .1005	— .0060	— .0910	— .1368	— .1316	— .0563
R	Raw	.4853	.5321	.4187	.4595	.4971	.4122
	Redc.	.4484	.4907	.4034	.4306	.4290	.3080

TABLE XVII  
ESTIMATED PER CENT OF VARIANCE OF THE GRADE SCORES  
ATTRIBUTABLE TO EACH TEST VARIABLE AND TO  
THE TOTAL THORNDIKE EXAMINATION

<i>Variable</i>	<i>Scores</i>	<i>Total G.</i>	<i>Con. Civ.</i>	<i>Sci. Math.</i>	<i>English</i>	<i>Soc. Sci.</i>	<i>For. Lang.</i>
1. Reading C.	Raw	3.9	13.9	— 0.1	7.2	5.1	1.2
	Redc.	5.0	13.8	0.9	5.7	4.6	1.8
2. Verbal Ability	Raw	9.0	6.9	5.2	12.1	10.2	7.0
	Redc.	7.6	8.2	3.3	11.6	9.4	6.1
I. Reading + Verbal	Raw	12.9	20.8	5.1	19.3	15.4	8.2
	Redc.	12.6	22.0	4.2	17.3	14.0	7.9
II. Math. Ability	Raw	5.9	3.1	9.1	2.5	4.2	1.5
	Redc.	4.5	2.7	7.1	2.6	3.5	0.9
III. Directions	Raw	3.1	— 0.1	3.7	— 0.2	1.2	1.1
	Redc.	4.3	— 0.5	5.8	— 0.1	2.1	1.3
IV. Information	Raw	1.7	4.5	— 0.4	— 0.5	3.9	6.2
	Redc.	— 1.3	— 0.1	— 0.9	— 1.2	— 1.2	— 0.5
Total Examination	Raw	23.6	28.3	17.5	21.1	24.7	17.0
	Redc.	20.1	24.1	16.2	18.6	18.4	9.6

Basing the interpretation primarily on the coefficients obtained from reduced scores, the results presented in Tables XVI and XVII may be summarized as follows:

1. The per cent of variance of the various grade criteria that may be attributed to the Thorndike Examination ranges from 9.5 to 24.0 per cent. This indicates a very low efficiency for the examination in predicting these individuals three-year grade scores.

2. Twenty-two per cent of the variance of the Contemporary Civilization grades is attributable to variable I; of this variable, fourteen per cent of the variance of these grade scores is attributable to the Reading Comprehension tests and eight per cent to the Verbal Ability tests; whereas only twenty-four per cent of the variance of these grade scores is attributable to the total examination. Thus, practically all of the variance of the Contemporary Civilization grades attributable to the total examination may be attributed directly to variable I.

3. Seventeen per cent of the variance of the English grades is attributable to variable I, nearly twelve per cent being attributable to the Verbal Ability tests alone; whereas only 1.5 per cent more, 18.5 per cent in all, may be attributed to the total examination.

4. Eight per cent of the variance of the Foreign Language grades is attributable to variable I, six per cent being attributable to the Verbal Ability tests alone; whereas only 9.5 per cent may be attributed to the total examination.

5. Fourteen per cent of the variance of the Social Science grades is attributable to variable I, nine per cent being attributable to the Verbal Ability tests alone; whereas only eighteen per cent may be attributed to the total examination.

6. About thirteen per cent of the variance of the total grade scores is attributable to variable I, about eight per cent being attributable to the Verbal Ability tests alone; whereas only twenty per cent may be attributed to the total examination.

7. Four per cent of the variance of the Science and Mathematics grades is attributable to variable I; whereas seven per cent is attributable to variable II, the Mathematical Ability tests; and sixteen per cent to the total examination.

8. Variables III and IV, the Following Directions and the General Information tests, have relatively little weight in de-

termining the variance of the grade criteria. Nearly six per cent of the variance of the Science and Mathematics grades, however, may be attributed to variable III, whereas only seven per cent is attributable to variable II, the Mathematical Ability tests.

9. On the whole, the per cent of variance of the grade scores attributable to variables II, III, or IV is relatively low; whereas, on the whole, the per cent of the variance of the grade scores attributable to variable I is nearly as great as the per cent attributable to the total examination. The principal exception to this estimate is in the case of the Science and Mathematics grades.

These estimates lend additional weight to the probable truth of the contention that the examination is measuring a common intellectual function that may best be characterized as verbal ability plus certain factors dependent upon the testing situation. Particularly, do these results support the view that the correlative relationships found to exist between the total examination (seventy-five per cent of its variance being estimated as attributable to the common factor, and only ten per cent to the specific factor) and the various grade criteria may be attributed to the verbal ability aspect of the common factor, since so much of the variance of the grade criteria attributable to the total examination is attributable to the Verbal Ability tests, *and*, since verbal ability is a common function of all of these college studies.

With the exception of the Science and Mathematics grades, most of the variance attributable to the total examination not attributable to the Verbal Ability tests may be attributed to the Reading Comprehension tests, which may be thought of as measuring, relatively, a considerable amount of verbal ability. That the group of Mathematical Ability tests is probably measuring number ability is indicated by its relationships with the Science and Mathematics grades; but that it is measuring less of number ability and more of verbal ability plus factors dependent upon the testing situation is very probable,—such an inference having been made from the estimates of its per cent of variance attributable to the specific and common factors.

## V. SUMMARY AND CONCLUDING INTERPRETATION

This investigation represents an analysis of the Thorndike Intelligence Examination for High School Graduates, an investigation made in order to determine an answer to the general question of what the examination measures and how adequately it measures it, as well as to determine an answer to the question: Is the Thorndike Examination a valid or adequately valid measure of *general scholastic ability*?

1. The problem was attacked by means of an analysis of the Thorndike Examination records of 568 male subjects, candidates for admission to Columbia College, taking the examination in June, 1925.

2. The battery of tests set up from the total examination satisfied the criterion for a common factor and specific factors when the Reading Comprehension tests were combined with the Verbal Ability tests.

3. The estimates were made that seventy-five per cent of the variance of the total examination could be attributed to the common factor, ten per cent to the specific factor, and fifteen per cent to chance errors of measurement.

4. By means of (1) the estimates of the per cent of variance of each group of tests attributable to the common, specific, and chance factors, (2) the estimates of the per cent of variance of the grade scores (for 159 subjects of the original group of 568, who entered Columbia College and for whom three-year grade scores were available) attributable to the various groups of tests and to the total examination, and (3) other criteria, such as the contents of the tests, factors a function of the testing situation, and various correlative relationships, the general conclusion was made that the common function so extensively measured by the examination might best be characterized as *verbal ability plus* certain factors dependent upon the testing situation.

5. Whether the common functions measured by the examination can be characterized as *general scholastic ability* is very doubtful, unless by general scholastic ability is meant nothing more than verbal ability plus factors dependent upon the testing situation.

6. Since scholastic records represent perhaps the best available validity criteria for an examination purporting to measure general scholastic ability, and in view of the estimates that

only 9.5 to 24.0 per cent of the variance of the various grade criteria used in this study is possibly attributable to the performance of these individuals on the total examination, it is very improbable that the concept *general scholastic ability* can be adequately defined in terms of verbal ability plus factors dependent upon the testing situation. This represents most probably only a partial definition.

7. The suggestion is offered that the uses made of the individual's test score on this examination could be better served by several scores for each individual derived from his performance on various groups of tests, each group designed to measure differential group functions, such as verbal ability, number ability, and memory ability, as well as groups of tests designed to measure achievement in various academic subject matters, such as physics, English grammar, history, German language, etc.

Since some of those individuals concerned with the administration and use of entrance examinations might consider such a testing program as too laborious, as fantastic from their point of view, the writer makes the further suggestion: if any judgments of an individual's intellectual fitness or unfitness to carry college work are at all seriously taken as a function of his psychological examination performance, such judgments are sufficiently important determinations, particularly with regard to the welfare of the candidates for admission, to warrant pragmatic estimates of the validity of a testing program, similar to that suggested here. For those administrators whose judgments are now served by scores on groups of achievement or placement tests as well as by the scores on the Thorndike Entrance Examination, the problem of estimating the validity of a testing program of this kind would probably be simplified (depending upon the scope, adequacy, and use of the achievement examinations) to an investigation of the validity of groups of tests designed to measure differential group functions, such as verbal ability, number ability, memory ability, and other groups of functions that may from time to time be revealed as fairly independent of each other and relevant to the indices of judgment desired.

8. The author wishes to emphasize that the Thorndike scores used in this study were derived from a rather highly selected group of high school graduates and that the analysis herein made is accordingly most applicable to a population of which this group is a sample.

## BIBLIOGRAPHY

1. Anastasi, A. A Group Factor in Immediate Memory. *Arch. Psychol.*, 1930, No. 120.
2. Asher, E. J. The Predictive Value of Mental Tests that Satisfy Spearman's Tetrad Criterion. *J. Appl. Psychol.*, 1929, 13, 152-158.
3. Cleeton, G. U. The Predictive Value of Certain Measures of Ability in College Freshmen. *J. Educ. Res.*, 1927, 15, 357-370.
4. Cowdery, K. M. Repeated Thorndike Intelligence Examinations. *School & Soc.*, 1928, 27, 367-369.
5. Crawford, A. B. Forecasting Freshman Achievement. *School & Soc.*, 1930, 31, 125-132.
6. Edgerton, H. A. Academic Prognosis in the University. Balt., Warwick, 1930.
7. Ezekiel, M. Methods of Correlation Analysis. N. Y., Wiley, 1930.
8. Flemming, E. G. The Predictive Value of Certain Tests of Emotional Stability as Applied to College Freshmen. *Arch. Psychol.*, 1928, No. 96.
9. Freeman, F. S. Elusive Factors Tending to Reduce Correlations between Intelligence Test Ranks and College Grades. *School & Soc.*, 1929, 29, 784-786.
10. Garrett, H. E. Statistics in Psychology and Education. N. Y., Long., 1926.
11. Garrett, H. E. A Study of the CAVD Intelligence Examination. *J. Educ. Res.*, 1930, 21, 103-108.
12. Grauer, D. and Root, W. T. The Thorndike Intelligence Tests and Academic Grades. *J. Appl. Psychol.*, 1927, 11, 297-318.
13. Guiler, W. S. The Predictive Value of Group Intelligence Tests. *J. Educ. Res.*, 1927, 16, 365-374.
14. Husband, R. W. The Reliability of the Thorndike Intelligence Examination. *School & Soc.*, 1928, 28, 521-522.
15. Jones, A. L. A Report of Professor Adam Leroy Jones' Speech to the American Institute of Mining and Metallurgical Engineers. *School & Soc.*, 1921, 17, 268-269.
16. Kelley, T. L. The Principles and Techniques of Mental Measurement. *Amer. J. Psychol.*, 1923, 34, 408-432.
17. Kelley, T. L. Statistical Method. N. Y., Macm., 1924.
18. Kelley, T. L. Crossroads in the Mind of Man. Stanford Univ., Stanford Univ. Press, 1928.
19. Lauer, A. R. and Evans, J. E. The Relative Predictive Value of Different High School Subjects on College Grades. *School & Soc.*, 1930, 31, 159-160.
20. MacPhail, A. H. The Intelligence of College Students. Balt., Warwick, 1924.
21. May, M. A. Predicting Academic Success. *J. Educ. Psychol.*, 1923, 14, 429-440.
22. Pearson, K. and Moul, M. The Mathematics of Intelligence. I. The Sampling Errors in the Theory of a Generalized Factor. *Biometrika*, 1927, 19, 246-291.
23. Pintner, R. Intelligence Testing. Rev. Ed. N. Y., Holt, 1931.
24. Schneek, M. R. The Measurement of Verbal and Numerical Abilities. *Arch. Psychol.*, 1929, No. 107.
25. Spearman, C. Correlations of Sums and Differences. *Brit. J. Psychol.*, 1912-13, 5, 417-426.
26. Spearman, C. The Abilities of Man. N. Y., Macm., 1927.

27. Spearman, C. and Holzinger, K. The Average Value for the Probable Error of Tetrad Differences. *Brit. J. Psychol.*, 1930, 20, 368-370.
28. Thomson, G. H. The Tetrad Difference Criterion. *Brit. J. Psychol.*, 1927, 17, 235-255.
29. Thorndike, E. L. Intelligence Examinations for College Entrance. *J. Educ. Res.*, 1920, 1, 329-337.
30. Thorndike, E. L. *et al.* The Measurement of Intelligence. N. Y., T. C., Col. Univ., 1927.
31. Thurstone, L. I. Psychological Examinations for College Freshmen. *J. Educ. Res.*, 1927, 1, 1-27.
32. Warren, R. and Mendenhall, R. M. The Mendenhall-Warren-Hollerith Correlation Method. N. Y., Col. Univ. Stat. Bur., Document No. 1, 1929.
33. Whipple, G. M. Intelligence Tests in Colleges and Universities. Twenty-first Yearbook of the National Society for the Study of Education, Ch. 10. Bloomington, Ill., Pub. School Publ. Co., 1922.
34. Wilson, J. H. The Nature of Intelligence. *J. Educ. Psychol.*, 1931, 22, 20-34.
35. Wood, B. D. Measurement in Higher Education. Yonkers, N. Y., World B. Co., 1923.
36. Yule, G. U. An Introduction to the Theory of Statistics. Lond., Griffin 1919.

150.8

A673

no. 128

1931

Archives of psychology

150.8

A673

no. 128

1931

Archives of psychology

*List of numbers, continued from inside front cover*

63. Race Differences in Inhibition: ALBERT L. CRANE. \$1.50.
64. Individual Differences in Incidental Memory: SADIE MYERS SHELLOW. \$1.25.
65. Character Traits as Factors in Intelligence Test Performance: WILLIAM M. BROWN. \$1.25.
66. A Study of the Sexual Interest of Young Women: F. I. DAVENPORT. \$1.25.
67. The Psychology of Confidence: WM. CLARK TROW. \$1.25.
68. Experimental Studies of College Teaching: HAROLD E. JONES. \$1.25.
69. The Influence of Treatment for Intestinal Toxemia on Mental and Motor Efficiency: ALICE E. PAULSEN. \$1.00.
70. A Study of Suggestibility of Children: MARGARET OTIS. \$1.50.
71. The Value of Praise and Reproof as Incentives for Children: ELIZABETH B. HURLOCK. \$1.00.
72. Attention and Interest in Advertising: HOWARD K. NIXON. \$1.25.
73. An Experimental Study of Thinking: EDNA HEDBREDE. \$1.75.
74. Estimation of Time: ROBERT AXEL. \$1.00.
77. Tested Mentality as Related to Success in Skilled Trade Training: THEODORA M. ABEL. \$1.25.
78. Aggressive Behavior in a Small Social Group: E. M. RIDDLE. \$1.75.
79. The Memory Value of Advertisements: EDITH R. BRANDT. \$1.25.
80. A Critical Examination of Test-Scoring Methods: ROSE G. ANDERSON. \$1.00.
81. Thermal Discrimination and Weber's Law: ELMER A. K. CULLER. \$1.75.
82. A Correlational Analysis of Typing Proficiency: LUTON ACKERSON. \$1.50.
83. Recall as a Function of Perceived Relations: CORA B. KEY. \$1.25.
84. A Study of the Consistency of Rate of Work: CONSTANCE E. DOWD. \$1.00.
85. An Experimental Investigation of Recovery from Work: S. L. CRAWLEY. \$1.25.
86. Facilitation and Inhibition: THOMAS N. JENKINS. \$1.00.
87. Variability of Performance in the Curve of Work: JAMES D. WEINLAND. \$1.00.
88. A Mental Hygiene Inventory: S. DANIEL HOUSE. \$1.50.
89. Mental Set and Shift: ARTHUR T. JERSILD. \$1.25.
90. An Experimental Investigation of Rest Pauses: CHARLES W. MANZER. \$1.25.
91. Routine and Varying Practice as Preparation for Adjustment to a New Situation: LELAND W. CRAFTS. \$1.00.
92. The Development of a Standardized Animal Maze: L. H. WARNER and G. J. WARDEN. \$1.25.
93. An Experimental Study of Speed and Other Factors in "Racial" Differences: OTTO KLINEBERG. \$1.50.
94. The Relation of Reaction Time to Measures of Intelligence, Memory, and Learning: VERNON W. LEMMON. 80 cents.
95. Is the Latent Time in the Achilles Tendon Reflex a Criterion of Speed in Mental Reactions? GEORGE H. ROUNDS. \$1.25.
96. The Predictive Value of Certain Tests of Emotional Stability as Applied to College Freshmen: EDWIN G. FLEMMING. \$1.00.
97. A Vocabulary Information Test: ANOELINA L. WEKS. \$1.00.
98. The Effect of Various Temporal Arrangements of Practice on the Mastery of an Animal Maze of Moderate Complexity: SIDNEY A. COOK. 80 cents.
99. Recognition Time as a Measure of Confidence: GEORGE HOFFMAN SEWARD. \$1.00.
100. Precision and Accuracy: GEORGE W. HARTMAN. 80 cents.
101. A Group Test of Home Environment: EDITH MARIE BURDICK. \$1.50.
102. The Effect of Changed Material on Ability to Do Formal Syllogistic Reasoning: MINNA CHEVES WILKINS. \$1.25.
103. The Effect of Incentives on Accuracy of Discrimination Measured on the Galton Bar: HUGHBERT C. HAMILTON. \$1.25.
104. An Investigation into the Validity of Norms with Special Reference to Urban and Rural Groups: MYRA E. SHIMBERG. \$1.25.
105. Blood Pressure Changes in Deception: MATTHEW N. CHAPPELL. 80 cents.
106. An Experimental Comparison of Psychophysical Methods: W. N. KELLOGG. \$1.25.
107. The Measurement of Verbal and Numerical Abilities: MATTHEW M. R. SCHNECK. \$1.00.
108. A Perseverative Tendency in Pre-School Children. A Study in Personality Differences: HAZEL MORTON CRISHING. \$1.00.
109. A Preliminary Study of the Effect of Training in Junior High School Shop Courses: L. DEWEY ANDERSON. 80 cents.
110. Music Appreciation: An Experimental Approach to its Measurement: MORTIMER JEROME ADLER. \$1.50.
111. Motivation in Fashion: ELIZABETH B. HURLOCK. \$1.00.
112. An Experimental Evaluation of Equality Judgments in Psychophysics: W. N. KELLOGG. \$1.00.
113. Illusions in the Perception of Short Time Intervals: NATHAN ISRAELI. \$.80.
114. Further Studies of the Reading-Recitation Process in Learning: E. B. SKAGGS, S. GROSSMAN, LOUISE KRUEGER, WM. C. F. KRUEGER. \$.80.
115. Factors Affecting the Galvanic Reflex: ROLAND C. DAVIS. \$1.00.
116. A Study of the Infant's Feeding Reactions During the First Six Months of Life: ROWENA RIPIN. 80 cents.
117. An Experiment in the Measurement of Mental Deterioration: HARRIET BABCOCK. \$1.25.
118. An Experimental Investigation of the Phenomenon of Postural Persistence: LOWELL SINN SELLING. \$1.00.
119. The American Council on Education Rating Scale: Its Reliability, Validity, and Use: FRANCIS FOSTER BRADSHAW. \$1.00.
120. A Group Factor in Immediate Memory: ANNE ANASTASI. \$1.00.
121. Individual Differences in the Sense of Humor and their Relation to Temperamental Differences: POLYXENIE KAMBOUROPOULOU. \$1.00.
122. Suggestibility in the Normal and Hypnotic States: GRIFFITH W. WILLIAMS. \$1.00.
123. An Analytical Study of the Conditioned Knee-Jerk: GEORGE RICHARD WENDT. \$1.25.
124. Race Differences in the Organization of Numerical and Verbal Abilities: JACK W. DUNLAP. \$1.25.
125. Errors of Measurement and Correlation: EDWARD E. CURETON. \$1.25.
126. Relationships between experience factors, test scores and efficiency as shown by a study of four groups of women office workers: NORMA BIRD. \$1.00.
127. Individual differences in delayed reactions of infants: a study of sex differences in early retentiveness: CHANCEY N. ALLEN. \$.80.
128. A Study of Factors Measured by the Thorndike Intelligence Examination for High School Graduates. \$1.00.

## DIRECTORY OF AMERICAN PSYCHOLOGICAL PERIODICALS

- AMERICAN JOURNAL OF PSYCHOLOGY**—Ithaca, N. Y.; Cornell University.  
 Subscription \$6.50. 624 pages annually. Edited by M. F. Washburn, K. M. Dallenbach, Madison Bentley, and E. G. Boring. Quarterly. General and experimental psychology. Founded 1887.
- JOURNAL OF GENETIC PSYCHOLOGY**—Worcester, Mass.; Clark University Press.  
 Subscription \$7.00. 600 pp. annually. Edited by Carl Murchison. Quarterly. Child behavior, animal behavior, comparative psychology. Founded 1891.
- PSYCHOLOGICAL REVIEW**—Princeton, N. J.; Psychological Review Company.  
 Subscription \$5.50. 540 pages annually. Edited by Howard O. Warren. Bi-monthly. General psychology. Founded 1894.
- PSYCHOLOGICAL MONOGRAPHS**—Princeton, N. J.; Psychological Review Company.  
 Subscription \$6.00 per vol. 500 pages. Edited by Raymond Dodge. Without fixed dates, each issue one or more researches. Founded 1895.
- PSYCHOLOGICAL INDEX**—Princeton, N. J.; Psychological Review Company.  
 Subscription \$4.00. 300-400 pages. Edited by Walter S. Hunter. An annual bibliography of psychological literature. Founded 1895.
- PSYCHOLOGICAL BULLETIN**—Princeton, N. J.; Psychological Review Company.  
 Subscription \$6.00. 720 pages annually. Edited by Edward S. Robinson. Monthly (10 numbers). Psychological literature. Founded 1904.
- ARCHIVES OF PSYCHOLOGY**—New York, N. Y.; Columbia University.  
 Subscription \$6.00. 500 pages per volume. Edited by R. S. Woodworth. Without fixed dates, each number a single experimental study. Founded 1906.
- JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY**—Eno Hall, Princeton, N. J.; American Psychological Association. Subscription \$5.00. 448 pages annually. Edited by Henry T. Moore. Quarterly. Abnormal and social. Founded 1906.
- PSYCHOLOGICAL CLINIC**—Philadelphia, Pa.; Psychological Clinic Press.  
 Subscription \$3.00. 288 pages. Edited by Lightner Witmer. Without fixed dates (9 numbers). Orthogenics, psychology, hygiene. Founded 1907.
- PSYCHOANALYTIC REVIEW**—Washington, D. C.; 3617 10th St., N. W.  
 Subscription \$6.00. 500 pages annually. Edited by W. A. White and S. E. Jelliffe. Quarterly. Psychoanalysis. Founded 1913.
- JOURNAL OF EXPERIMENTAL PSYCHOLOGY**—Princeton, N. J.; Psychological Review Company.  
 Subscription \$6.00. 500 pages annually. Edited by Samuel W. Fernberger. Bi-monthly. Experimental psychology. Founded 1916.
- JOURNAL OF APPLIED PSYCHOLOGY**—Baltimore, Md.; Williams & Wilkins Company.  
 Subscription \$5.50. 400 pages annually. Edited by James P. Porter. Bi-monthly. Founded 1917.
- JOURNAL OF COMPARATIVE PSYCHOLOGY**—Baltimore, Md.; Williams & Wilkins Company. Subscription \$5.00 per volume of 450 pages. Three volumes every two years. Edited by Knight Dunlap and Robert M. Yerkes. Founded 1921.
- COMPARATIVE PSYCHOLOGY MONOGRAPHS**—Baltimore, Md.; The Johns Hopkins Press.  
 Subscription \$5.00. 400 pages per volume. Knight Dunlap, Managing Editor. Published without fixed dates, each number a single research. Founded 1922.
- GENETIC PSYCHOLOGY MONOGRAPHS**—Worcester, Mass.; Clark University Press. Subscription \$7.00 per volume of 500-600 pages. Two volumes per year. Edited by Carl Murchison. Monthly. Each number one complete research. Child behavior, animal behavior, and comparative psychology. Founded 1925.
- PSYCHOLOGICAL ABSTRACTS**—Eno Hall, Princeton, N. J.; American Psychological Association.  
 Subscription \$6.00. 700 pages annually. Edited by Walter S. Hunter. Monthly. Abstracts of psychological literature. Founded 1927.
- JOURNAL OF GENERAL PSYCHOLOGY**—Worcester, Mass.; Clark University Press.  
 Subscription \$7.00. 500-600 pages annually. Edited by Carl Murchison. Quarterly. Experimental, theoretical, clinical, historical psychology. Founded 1927.
- JOURNAL OF SOCIAL PSYCHOLOGY**—Worcester, Mass.; Clark University Press.  
 Subscription \$7.00. 500-600 pages annually. Edited by John Dewey and Carl Murchison. Quarterly. Political, racial, and differential psychology. Founded 1929.